Article
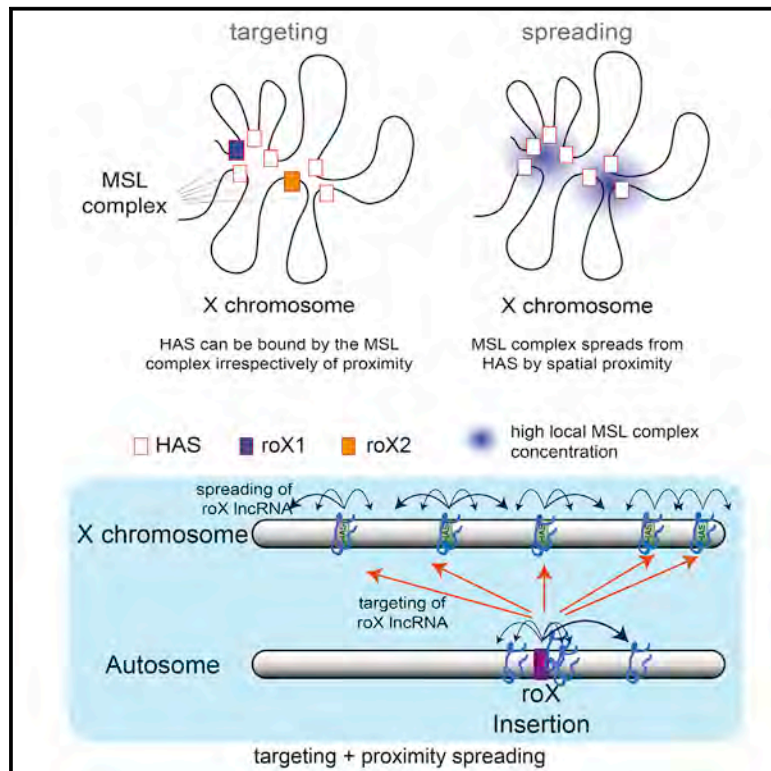
High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X Chromosome in *Drosophila*

Graphical Abstract

Authors
Fidel Ramírez, Thomas Lingg, Sarah Toscano, ..., Job Dekker, Thomas Manke, Asifa Akhtar

Correspondence
akhtar@ie-freiburg.mpg.de

In Brief
Combining chromosome conformation analyses with fly genetics, Ramírez et al. show that high-affinity sites (HAS), positioned at hubs of long-range chromatin contacts, arrange in a sex- and MSL complex-independent manner on the X chromosome. The MSL complex spreads via spatial proximity and regulates local chromatin remodeling rather than influencing global chromatin architecture.

Highlights

- HAS frequently occur at regions with enriched long-range contacts on the X chromosome

- Global X chromosome architecture is sex- and MSL complex-independent in flies

- roX HAS dynamically organize within the X territory depending on transcriptional status

- The MSL complex uses spatial proximity to spread and affects nucleosome pattern at HAS

Accession Numbers
GSE58821

CrossMark

CellPress

High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X Chromosome in *Drosophila*

Fidel Ramírez,[1,7] Thomas Lingg,[1,2,7] Sarah Toscano,[1,7] Kin Chung Lam,[1,2,7] Plamen Georgiev,[1] Ho-Ryun Chung,[3] Bryan R. Lajoie,[4] Elzo de Wit,[5] Ye Zhan,[4] Wouter de Laat,[5] Job Dekker,[4,6] Thomas Manke,[1] and Asifa Akhtar[1,*]

[1]Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany
[2]Faculty of Biology, University of Freiburg, 79104 Freiburg, Germany
[3]Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[4]Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605-0103, USA
[5]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands
[6]Howard Hughes Medical Institute
[7]Co-first author
*Correspondence: akhtar@ie-freiburg.mpg.de
http://dx.doi.org/10.1016/j.molcel.2015.08.024

SUMMARY

Dosage compensation mechanisms provide a paradigm to study the contribution of chromosomal conformation toward targeting and spreading of epigenetic regulators over a specific chromosome. By using Hi-C and 4C analyses, we show that high-affinity sites (HAS), landing platforms of the male-specific lethal (MSL) complex, are enriched around topologically associating domain (TAD) boundaries on the X chromosome and harbor more long-range contacts in a sex-independent manner. Ectopically expressed roX1 and roX2 RNAs target HAS on the X chromosome in *trans* and, via spatial proximity, induce spreading of the MSL complex in *cis*, leading to increased expression of neighboring autosomal genes. We show that the MSL complex regulates nucleosome positioning at HAS, therefore acting locally rather than influencing the overall chromosomal architecture. We propose that the sex-independent, three-dimensional conformation of the X chromosome poises it for exploitation by the MSL complex, thereby facilitating spreading in males.

INTRODUCTION

The organization of chromosomes within the nucleus and the spatial arrangement of genes within a chromosome territory are gaining fundamental importance during epigenetic control of gene expression (Quinodoz and Guttman, 2014). Notably, the regulatory mechanisms of sex chromosomes offer ideal paradigms to understand how the expression of an entire chromosome, and, therefore, thousands of genes at once, can be controlled by epigenetic mechanisms (Brockdorff and Turner, 2015).

Dimorphic sex chromosomes genetically determine sex in many organisms. In the XX/XY sex determination system, males are heterogametic (XY), and females are homogametic (XX). To overcome the risk of an unequal transcriptional output, different organisms have evolved independent strategies (termed "dosage compensation") to balance the X chromosomal gene dose between the sexes (Vicoso and Charlesworth, 2006). In mammals, expression of the long non-coding RNA (lncRNA) Xist from only one of the two female X chromosomes leads to recruitment of silencing complexes in *cis* through which this chromosome becomes compacted and heterochromatinized (Heard and Disteche, 2006). In *Drosophila melanogaster*, dosage compensation happens on the single male X chromosome by formation of the male-specific lethal (MSL) complex, which promotes an approximately 2-fold transcriptional upregulation (Conrad and Akhtar, 2011). The MSL complex consists of four core proteins (MSL1, MSL2, MSL3, and males absent on the first [MOF]) which, together, form a hetero-octameric complex that is further stabilized by the integration of two lncRNAs, called RNA on the X chromosome (roX) 1 and 2, by the ATP-dependent RNA helicase maleless (MLE) (Keller and Akhtar, 2015). The formation of this ribonucleoprotein complex is believed to occur at the *roX* gene locus because roX RNAs are the only components of the complex being produced within the nucleus.

Based on both genetic and genomic analyses, the complex is thought to first target genomic regions called high-affinity sites (HAS), which include the *roX* genes, and then to spread to lower-affinity sites. During this process, MOF acetylates histone H4 lysine 16 across the entire X chromosome, which ultimately upregulates transcription (Conrad and Akhtar, 2011). However, how HAS are organized to allow the complex to reach the whole X chromosome continues to be an enigma. The MSL complex preferentially binds to an active chromatin environment containing a consensus sequence motif, called the MSL recognition

A

B

C

D
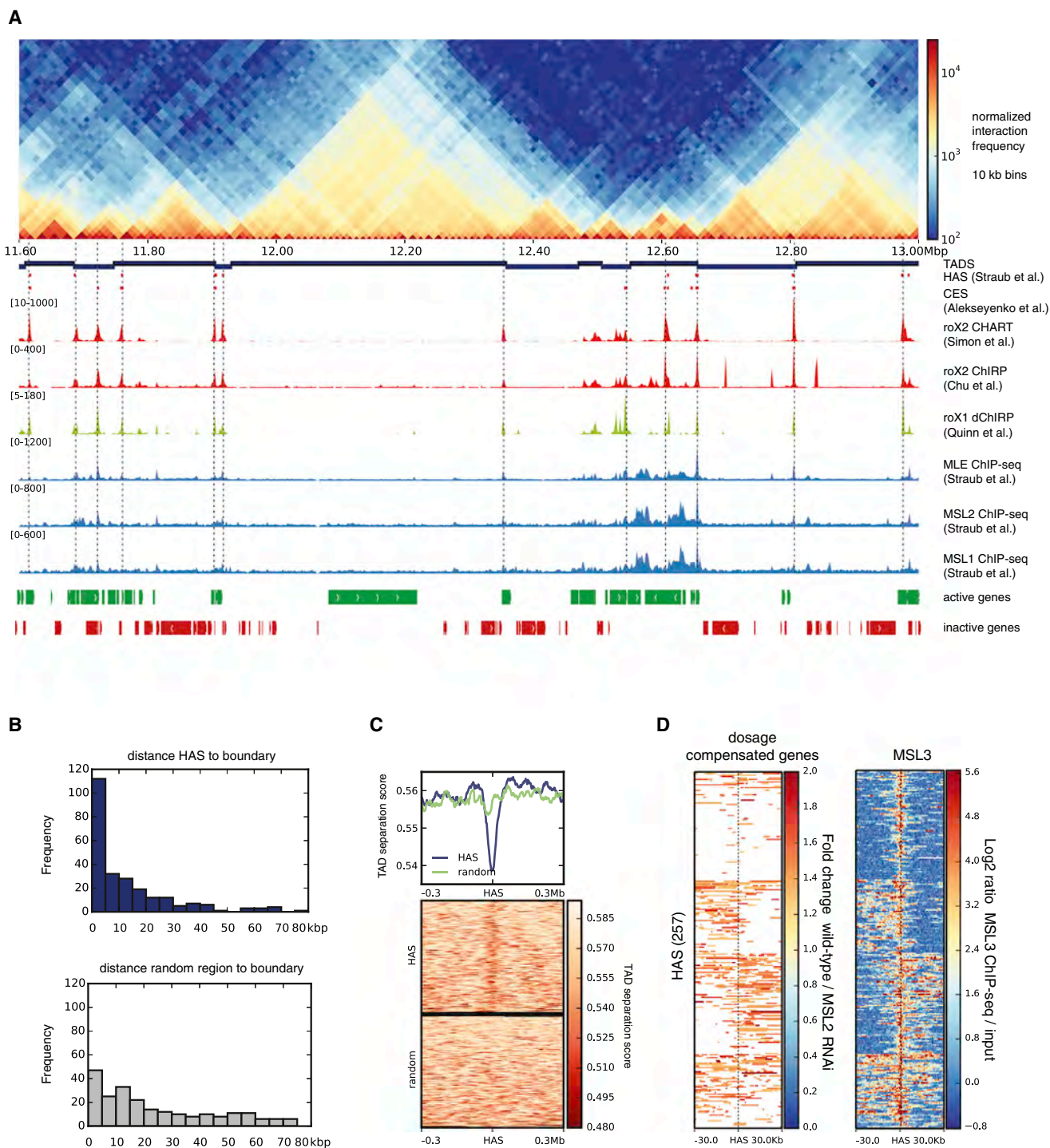
Figure 1. HAS Are Enriched at TAD Boundaries

(A) Normalized Hi-C counts at 10-kb resolution for the region 11.6–13.0 Mb in chromosome X of S2 cells. From the top, the tracks are as follows: partitioning of the genome into TADs; HAS as defined by Straub et al. (2008); HAS reported by Alekseyenko et al. (2008) (originally called chromosome entry sites [CESs]); roX2 CHART (Simon et al., 2011); roX2 ChIRP (Chu et al., 2011); roX1 domain-specific ChIRP (Quinn et al., 2014); MLE, MSL1, and MSL2 ChIP-seq (Straub et al., 2013); and active and inactive genes in S2 cells (Cherbas et al., 2011). Vertical lines are high-resolution HAS based on roX2 and MSL2 binding.

(B) Distribution of distances from the boundaries to HAS (top, blue) and from boundaries to the same number of shuffled random regions (bottom, gray) within chromosome X. The x axis represents the distance in kilobases from HAS to the boundary in bins of 5 kb. The y axis represents the number of HAS per bin.

(C) TAD separation score (Supplemental Experimental Procedures) around HAS showing the tendency of the MSL complex to land at boundaries. Lower scores indicate better TAD separation.

(legend continued on next page)

element (MRE), that is flanked by sequences of elevated GC content (Alekseyenko et al., 2012; Conrad and Akhtar, 2011). However, these features, although moderately enriched on the X chromosome, are also found on autosomes, and, therefore, it has not been possible to fully characterize HAS. Current models are based on linear genomic analysis or DNA fluorescent in situ hybridization (DNA FISH) on a few individual loci (Grimaud and Becker, 2009) without accounting for the potential influence of global chromosome conformation.

In this study, we used genome-wide chromosome conformation capture (Hi-C), a technique that enables the study of all chromosomal interactions within a genome at once (Lieberman-Aiden et al., 2009). This was further complemented by circularized chromosome conformation capture followed by deep sequencing (4C-seq) (Splinter et al., 2012) and three-dimensional double label DNA FISH (3D DNA FISH) analysis on single cells to study the interaction patterns of individual loci on the X chromosome. Our data highlight a distinct mechanism in flies in which specific features at topologically associating domain (TAD) boundaries on the X chromosome provide an advantageous location for the MSL complex to spread to spatially close regions and induce dosage compensation. Moreover, we show that, rather than modifying global chromosomal domain organization, the MSL complex acts locally by inducing chromatin remodeling at HAS.

RESULTS

High-Affinity Sites Occur Preferentially at Boundaries of Topologically Associating Domains

Previously published Hi-C studies in fruit flies could not address male-specific dosage compensation because either sex-mixed embryos (Sexton et al., 2012) or the female Kc cell line (Hou et al., 2012) alone were used. Therefore, we generated wild-type Hi-C contact maps using the restriction enzyme HindIII for two widely used male model cell lines, CME W1 cl.8+ (Currie et al., 1988) (clone-8) and Schneider's line 2 (Schneider, 1972) (S2) in biological duplicates. To ensure consistent comparisons, we also reprocessed previously published Hi-C contact maps for mixed-sex embryos and for Kc cell lines using the same mapping and normalization procedures (Figures S1A and S1B; Table S1). The correlation within replicates was very high for raw and corrected Hi-C counts (0.96 Pearson correlation in both cases, Figures S1C–S1F) as was the correlation between different cell types when raw and corrected Hi-C counts were considered. For the corrected counts we see an expected power-law decay of interaction frequencies with increasing genomic distance (Hou et al., 2012; Lieberman-Aiden et al., 2009; Sexton et al., 2012) and a similar decay for all chromosomes in the different Hi-C datasets (Figures S1G–S1I).

When combining our Hi-C data for S2 cells with published high-resolution roX occupancy sites (indicative of HAS) from capture hybridization analysis of RNA targets (CHART) (Simon et al., 2011) and chromatin isolation by RNA purification (ChIRP) (Chu
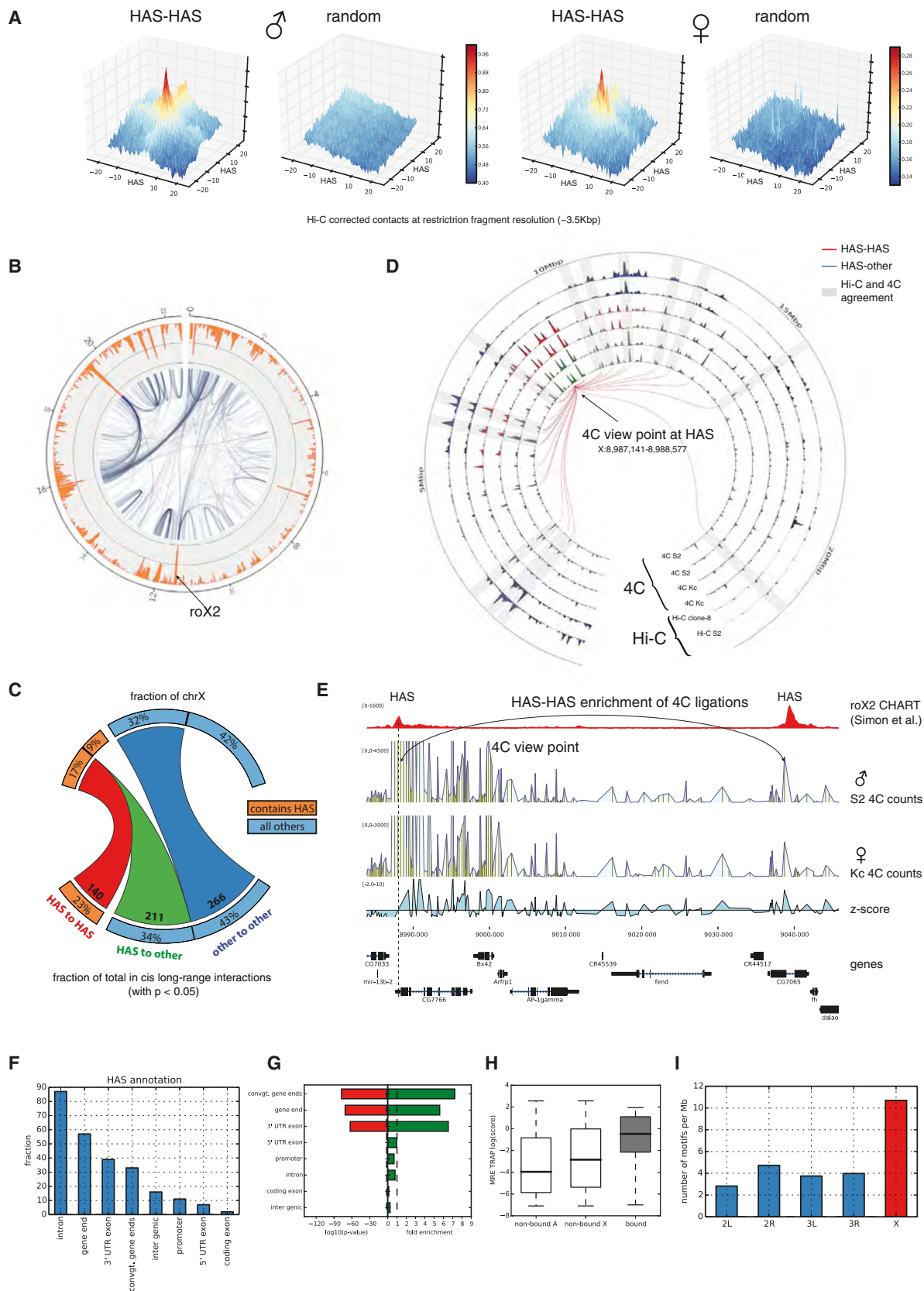
et al., 2011) for this cell type, we observed that HAS have a tendency to localize at or near TAD (Dixon et al., 2012; Hou et al., 2012; Nora et al., 2012; Sexton et al., 2012) boundaries on the X chromosome. This can be seen readily by simple visual inspection of the data (Figure 1A; Figures S2A and S2B). To corroborate this finding, we used a domain caller (Experimental Procedures; Figures S3A–S3C) to define TAD boundaries in S2 and clone-8 cells. Because the TAD structure is highly conserved between the cell types studied (Figure S3D), the boundaries obtained for S2 and clone-8 cells are very similar to the published domain partitions for Kc (Hou et al., 2012) and *Drosophila* embryos (Sexton et al., 2012; Figures S3E and S3F). For the X chromosome, we identified a total of 257 HAS using the genome-wide mapping of roX2 (Simon et al., 2011) and MSL2 (Straub et al., 2013) (see also Table S2 and Supplemental Experimental Procedures). This association to TAD boundaries is significantly different from a random distribution (p = 1.9×10^{-11}, Fisher's exact test, on the number of overlaps between boundaries and HAS; Supplemental Experimental Procedures). We found that 68% of HAS lie within a 20-kb distance of the nearest boundary. When calculated within a 5-kb distance, 45% of HAS are located near a boundary, in contrast to 9% expected by chance (Figure 1B; Figure S3G). To further validate this finding, we computed the TAD separation score at each HAS (Supplemental Experimental Procedures; Figures S3A and S3B) and verified that it tends to have a minimum, which is indicative of boundary regions, at HAS (Figure 1C). For comparison, similarly low values are obtained when the TAD separation score is evaluated at architectural protein binding sites (APBSs) (Van Bortle et al., 2014), which are thought to localize at TAD boundaries (Dixon et al., 2012; Hou et al., 2012; Sexton et al., 2012; Sofueva et al., 2013; Van Bortle et al., 2014; Figure S3H). When studying dosage-compensated genes (measured by downregulation of expression upon MSL2 depletion; Zhang et al., 2010), we found that they appear frequently either upstream or downstream of HAS (Figure 1D) in a pattern similar to the one observed for TAD boundaries that separate active and inactive chromatin (Figure S3I). Taken together, we conclude that the apparent linear arrangement of HAS along the genome follows a particular pattern dictated by the 3D TAD organization.

HAS Show More Enriched Hi-C Contacts in a Sex-Independent Manner

Previous analyses of Hi-C data have shown that active chromatin regions tend to interact (Lieberman-Aiden et al., 2009; Sexton et al., 2012). Moreover, boundary regions have been shown to be enriched in Hi-C contacts (Hou et al., 2012; Rao et al., 2014), and long-range contacts between architectural proteins have also been suggested (Liang et al., 2014). Because HAS are frequently associated with active chromatin (Alekseyenko et al., 2012) and appear at boundaries, we explored the possibility that the spreading in *cis* of the MSL complex could be mediated by long-range associations of HAS. Using a method similar to the paired end spatial chromatin analysis (PE-SCAn) (de Wit et al.,

---

(D) Dosage-compensated genes (based on RNA-seq data from wild-type and MSL2 RNAi-treated S2 cells; Zhang et al., 2010) and MSL3 ChIP-seq (Straub et al., 2013) up to 30 kb away from HAS. For dosage compensation, only genes showing activity are colored. The heatmap rows were divided into four clusters based on the kmeans algorithm using deepTools (Ramírez et al., 2014).
See also Tables S1 and S2 and Figures S1, S2, and S3.

A
HAS-HAS ♂ random   HAS-HAS ♀ random

Hi-C corrected contacts at restrictrion fragment resolution (~3.5Kbp)

B

roX2

D
— HAS-HAS
— HAS-other
Hi-C and 4C agreement

10Mbp

15Mbp

4C view point at HAS
X:8,987,141-8,988,577

5Mbp

20Mbp

4C S2
4C S2
4C Kc
4C Kc
4C

Hi-C clone-8
Hi-C S2
Hi-C

C
fraction of chrX
32%   42%
9%
17%
140
HAS to HAS
23%
211
HAS to other
34%
266
other to other
43%

contains HAS
all others

fraction of total in cis long-range interactions
(with p < 0.05)

E
HAS   HAS-HAS enrichment of 4C ligations   HAS
[0-1600]   roX2 CHART (Simon et al.)
[0.0-4500]   4C view point
♂ S2 4C counts
[0.0-3000]   ♀ Kc 4C counts
[-2.0-10]   z-score

8990.000   9000.000   9010.000   9020.000   9030.000   9040.000

genes
CG7033   CR45539   CR44517
mir-13b-2   Bx42   fend   CG7065
              Arfrp1   fh
       CG7766   AP-1gamma   dalao

F
HAS annotation

intron
gene end
3' UTR exon
convgt. gene ends
inter genic
promoter
5' UTR exon
coding exon

G
convgt. gene ends
gene end
3' UTR exon
5' UTR exon
promoter
intron
coding exon
inter genic

log10(p-value)   fold enrichment
-120 -90 -60 -30 0 1 2 3 4 5 6 7 8 9

H
MRE TRAP log(score)

non-bound A   non-bound X   bound

I
number of motifs per Mb

2L   2R   3L   3R   X

CellPress

2013; Figure 2A), we found that HAS, and, in general, TAD boundaries on all chromosomes, are frequently enriched for Hi-C contacts in different cell lines, including female Kc cells (Figure 2A; Figures S4A and S4B). We also investigated the differences of *cis* long-range contacts on the X chromosome for each cell type and compared them with each other using a rigorous estimation of significant long-range Hi-C contacts (Figure 2B; Figures S4C and S4D; Supplemental Experimental Procedures). To gain statistical confidence and to avoid spurious results, we segmented the genome into 25-kb bins for this analysis. Our results revealed that HAS tend to exhibit more long-range contacts than other regions. HAS were found in 26% of the 25-kb bins on the X chromosome, and these regions contribute to 57% of all significant intra-X chromosomal long-range Hi-C contacts (Figure 2C). The spatial association between different HAS is highly significant because only 77 interactions (11%) are expected by chance ($p = 5 \times 10^{-16}$, Fisher's exact test, one-tailed). Similar results were found for all studied cell types when using different p value thresholds to call long-range contacts and for boundaries on other chromosomes (Table S3). The same analysis using 10-kb bins from a merge of all Hi-C samples for S2 cells (Supplemental Experimental Procedures) show comparable results (Figure S4E; see also Figure S4F for examples of enriched HAS-HAS contacts). Because the Hi-C results represent population averages where each cell may contain a unique chromosomal conformation (Nagano et al., 2013), we do not expect that all enriched HAS contacts are found in each cell but, rather, that HAS form a dynamic interaction network on the X chromosome. We analyzed long-range contacts for bins containing either a HAS and a boundary, a HAS only, or a boundary only. We found that bins that contained a HAS (either alone or together with a boundary) had, on average, more long-range contacts than the bins that only contained a boundary (Figure S5A). We additionally compared bins containing a boundary on the X chromosome with those on autosomes and found the average number of enriched contacts to be similar. Although the majority of HAS were in the same bins as boundaries, all bins that contained a HAS had, on average, more long-range contacts as the bins that only contained a boundary but no HAS (Figure S5A). This result suggests that HAS tend to be associated to boundaries with more long-range contacts.

We validated our Hi-C results using 4C-seq on male S2 and female Kc cell lines (for 4C viewpoints, see Table S4). Figure 2D and Figure S5B show comparisons of adjusted p values where the high correspondence between Hi-C and 4C can be seen. These comparisons show that our estimations of long-range contacts are reliable and can be reproduced by a different experimental and processing method as the one used for Hi-C. Furthermore, the 4C data allowed us to explore the enriched contacts at short genomic distances (5–10 kb) that are not possible with Hi-C. At this resolution, we consistently detect HAS-HAS-enriched contacts that remained unchanged between the male and female cell lines (Figure 2E; Figure S5C).

Depletion of MSL2 or MSL3 Does Not Change Global TAD Organization

The similarity between male and female results from Hi-C and 4C experiments suggested that, contrary to expectations (Grimaud and Becker, 2009), the MSL complex may not alter the conformation of the X chromosome. To directly investigate this, we generated Hi-C data in male S2 cells depleted of the MSL complex members (via RNAi-mediated knockdown of either MSL2 or MSL3; Figure S6A) and compared it with control knockdown (EGFP RNAi) or wild-type Hi-C samples. The resulting Hi-C counts showed a high correlation between all samples (Figure S6B), whereas the HAS-HAS-enriched contacts of the knockdown samples did not differ from those of the wild-type or EGFP RNAi control (Figure S6C; Table S3), and the TAD structure remained virtually identical (Figure S6D). These data indicate that the dosage compensation machinery in *Drosophila* does not broadly alter chromosomal topology but, rather, acts over a pre-existing chromosome conformation independent of sex.

Figure 2. HAS Show More Hi-C and 4C Ligations with Other HAS

(A) Enrichment of Hi-C contacts between HAS loci in male (S2) and female (Kc) data (Experimental Procedures). The x and y axes show the distance from HAS in restriction fragment resolution. The z axis contains the mean value of all pooled normalized sub-matrices corresponding to a HAS-HAS intersection.

(B) Representation of long-range Hi-C contacts from S2 cells as a network (25-kb bins, p < 0.05; Experimental Procedures). The orange outer rim shows the total number of long-range Hi-C contacts per bin.

(C) Quantification of long-range interactions on chromosome X from HAS to HAS (red), from HAS to other regions (green), and within other regions (dark blue). Multiple HAS found in the same bins were considered once.

(D) Comparison of enriched 4C contacts in males and females and Hi-C contacts in two cell lines for the viewpoint at HAS position X:8,987,141–8,988,577. Tracks displayed from outside toward the inside are as follows: Hi-C S2- and Hi-C clone-8-adjusted –log(p values) for a 25-kb binning; 4C contact enrichment –log(p values) for Kc replicate 1, Kc replicate 2, S2 replicate 1, and S2 replicate 2 at 25-kb binning. Lines represent enriched contacts detected in one of the 4C S2 replicates. Red lines indicate interactions between HAS, and the blue line indicates interactions from HAS to other regions. Peripheral numbering indicates the genomic position in megabases. Grey rectangles highlight the agreement between Hi-C and 4C at the enriched 4C contacts at the connecting lines. Close to the viewpoint, the detection of enrichments that are close to TAD size is limited when using Hi-C data.

(E) 4C data for a viewpoint at the end of gene *CG7766* confirms the interaction between two HAS that are 50 kb apart. Further 4C examples are shown in Figure S5C.

(F) Graph representing the location of HAS on different regions in and around genes.

(G) Enrichment of MREs bound by the MSL complex in contrast to the genome-wide background occurrence of MREs.

(H) TRAP score for MREs on autosomes (non-bound A), MREs not bound by the MSL complex on chromosome X (non-bound X), and MREs bound to the MSL complex (bound).

(I) Occurrence of MREs at different chromosomes filtered by the following criteria. The MRE has to be on active chromatin, determined using H3K36me3 (GEO: GSM685608), the MRE has to be located near the gene end, the MRE needs to have a TRAP score of at least −2, and the MRE should be located within 25 kb of the nearest boundary. MREs on chromosome X are enriched for these features.

See also Tables S3 and S4 and Figures S4 and S5.

A

No. of active genes close to HAS

B

***Mann-Whitney rank
test p-value < 0.0001

C
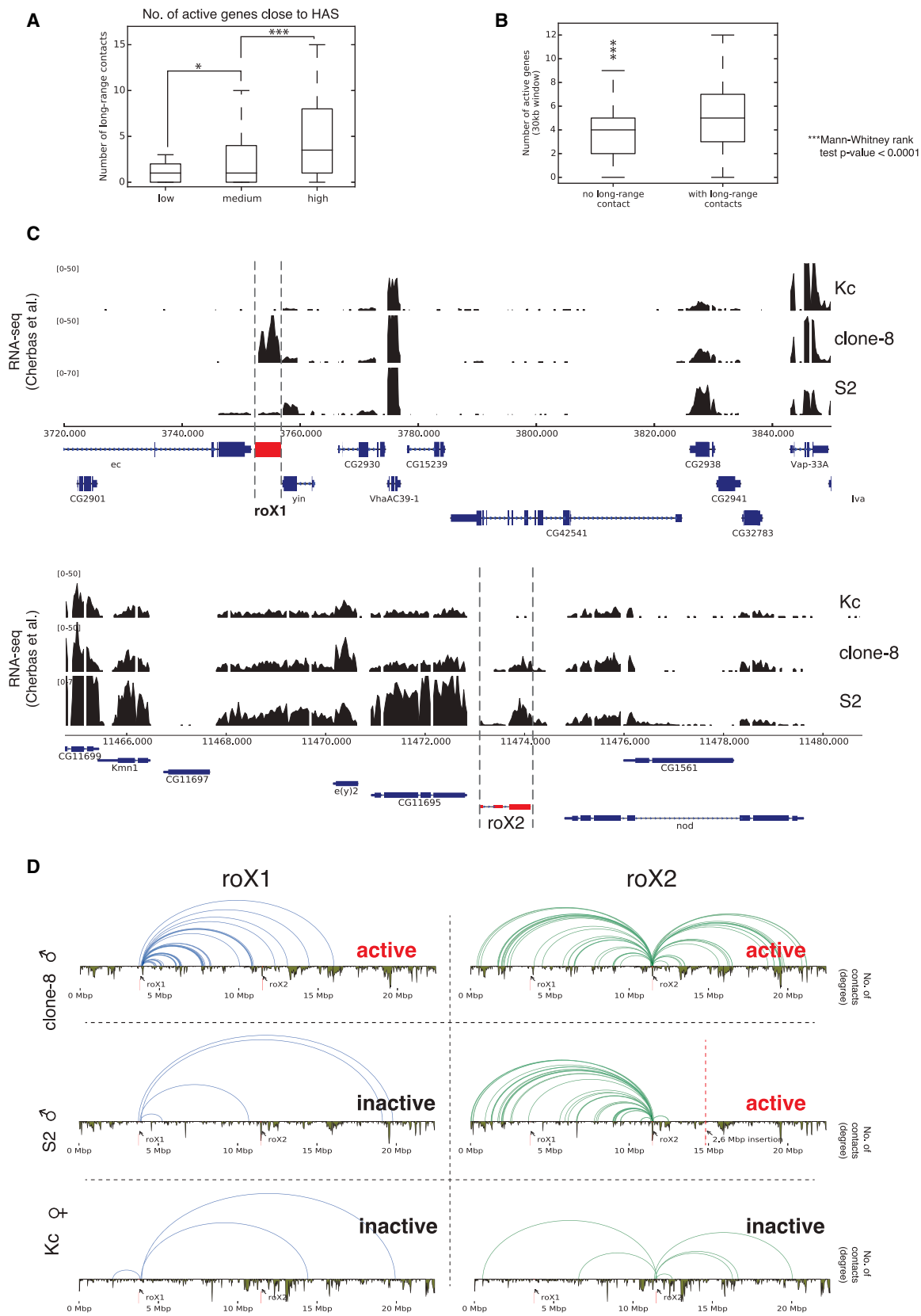
(legend on next page)

The X Chromosome Harbors Stronger MREs at Boundaries Compared with Autosomes

Because TAD boundaries are present on all chromosomes and are frequently enriched for long-range contacts (Figure S4B), we next addressed the relationship between TADs and HAS located on the X chromosome. Previous lower-resolution approaches have shown that HAS tend to be located at gene bodies or at the end of genes (Alekseyenko et al., 2008; Gilfillan et al., 2006). Using the high-resolution HAS derived from the CHART (Simon et al., 2011) and ChIRP (Chu et al., 2011) methods, we observe that HAS often appear on intronic regions (35%) and in the proximity of gene ends (51%), which include the 3′ UTR exon and convergent gene ends (Figure 2F). Moreover, HAS are almost never found at coding exons. The genomic distribution of the MRE associated to HAS over the different gene annotations revealed that gene ends are significantly enriched for MREs that are bound by the MSL complex (i.e., HAS) (Figure 2G). Also, HAS are always associated with active genes decorated with the histone H3 lysine 36 trimethylation (H3K36me3) histone mark and tend to have a DNA sequence with higher binding energy (transcription factor affinity prediction [TRAP] score; Thomas-Chollier et al., 2011), containing usually several copies of the MRE. When we considered only MREs that are at gene ends, are in active chromatin, have a log (TRAP score) of higher than -2, and are within 20 kb of the nearest boundary, we found that such a combination of features is enriched on chromosome X (Figures 2H and 2I). These data suggest that a combination of MREs, chromatin state, and gene architecture is required for the specificity of the MSL complex toward the X chromosome.

Differential Positioning of Active Regions within the X Chromosomal Territory

Correlation of expression and long-range contacts revealed that transcriptionally active HAS show more contacts compared with HAS located within inactive genes (Figures 3A and 3B). Analysis of two prominent HAS, roX1 and roX2, allowed us to explore this finding in more detail. In clone-8 cells, both roX RNAs are actively transcribed (Cherbas et al., 2011), whereas more than 99.5% of S2 cells do not show roX1 expression (Johansson et al., 2011), and female Kc cells do not express any of the roX genes (Cherbas et al., 2011). In contrast, genes surrounding roX1 and roX2 are expressed similarly in the three cell lines (Figure 3C). We observed a greater number of long-range contacts when the roX genes are active (Figure 3D; Figures S6E and S6F), although no changes were seen on the TAD structure. Next we explored whether this difference between the two

roX loci in the number of long-range contacts is reflected by their nuclear positioning in individual cells. For this, we performed 3D DNA FISH for roX1 and roX2 in clone-8 and S2R⁺ cells (a derivative of the original S2 cells with a similar transcriptional profile) and measured the radial distances of the probes to the center of mass of the MSL1-immunostained region (MSL territory) (Figure 4A; Table S5). In clone-8 cells, both expressed roX genes showed almost equal distance distributions, whereas, in S2R⁺ cells, the non-expressed roX1 appeared to be farther away compared with the expressed roX2 (p = 1.9 × 10⁻⁴, t test, one-tailed) (Figure 4B). In clone-8 cells, both roX probes can be found almost equally often outside of the MSL territory (17% of roX1 and 14% of roX2 signals), whereas, in S2R⁺ cells, roX1 is found outside more than three times as often as roX2 (28% versus 8%, Figure 4C). Taken together, the differences in the relative locations of roX1 and roX2 with respect to the MSL territory in clone-8 and S2R⁺ cells suggest that such positioning is related to their transcriptional activity. Consistent with this hypothesis, we observed that a FISH probe over a non-HAS (dpr8, inactive gene) was more frequently found outside of the MSL territory in comparison with two different HAS (HAS1 and HAS2, active genes) with respect to roX2 (Figure 4D). These data suggest that differential positioning of active regions within chromosomes could serve as an elegant mechanism for tissue-specific fine-tuning without changes in TAD structure, therefore providing plasticity for gene regulation while maintaining stability of the overall chromosome shape.

The MSL Complex Spreads from HAS to Spatially Proximal Regions

Knowing that HAS are located at regions often engaging in long-range contacts, we tested whether this is a sufficient condition for targeting of the MSL complex to loci in 3D proximity to the roX genes by studying a previously unnoticed, large (~2.67-Mb) insertion of chromosome 3L (chr3L, ~796,745–3,468,912) into the X chromosome (chrX, ~14,809,484) that was revealed by our Hi-C analysis of S2 cells (Figure S7A) and is most likely present in one of the two X chromosomes in the tetraploid S2 cells. This large region appears to be enriched for the MSL3 protein—a hallmark of MSL complex spreading in chromatin immunoprecipitation sequencing (ChIP-seq) data.—although we did not detect typical HAS. In Figure 5A, we only see seemingly spurious roX peaks that are not consistent between the two methods used to determine roX-DNA contacts (CHART, Simon et al., 2011; ChIRP, Chu et al., 2011) and did not observe other MSL proteins. For comparison, an X-linked region of similar size as the insertion contains, on average, 30 HAS. Using a

Figure 3. Comparison of Gene Expression Activity and Long-Range Contacts

(A) Comparison of the number of active genes associated to HAS and the number of long-range contacts. We stratified the number of active genes within a 15-kb window up- and downstream of HAS into low (<4 active genes), medium (between 4 and 7 active genes) and high (8 or more active genes). *p < 0.01, ***p < 0.0001, Mann-Whitney rank test.
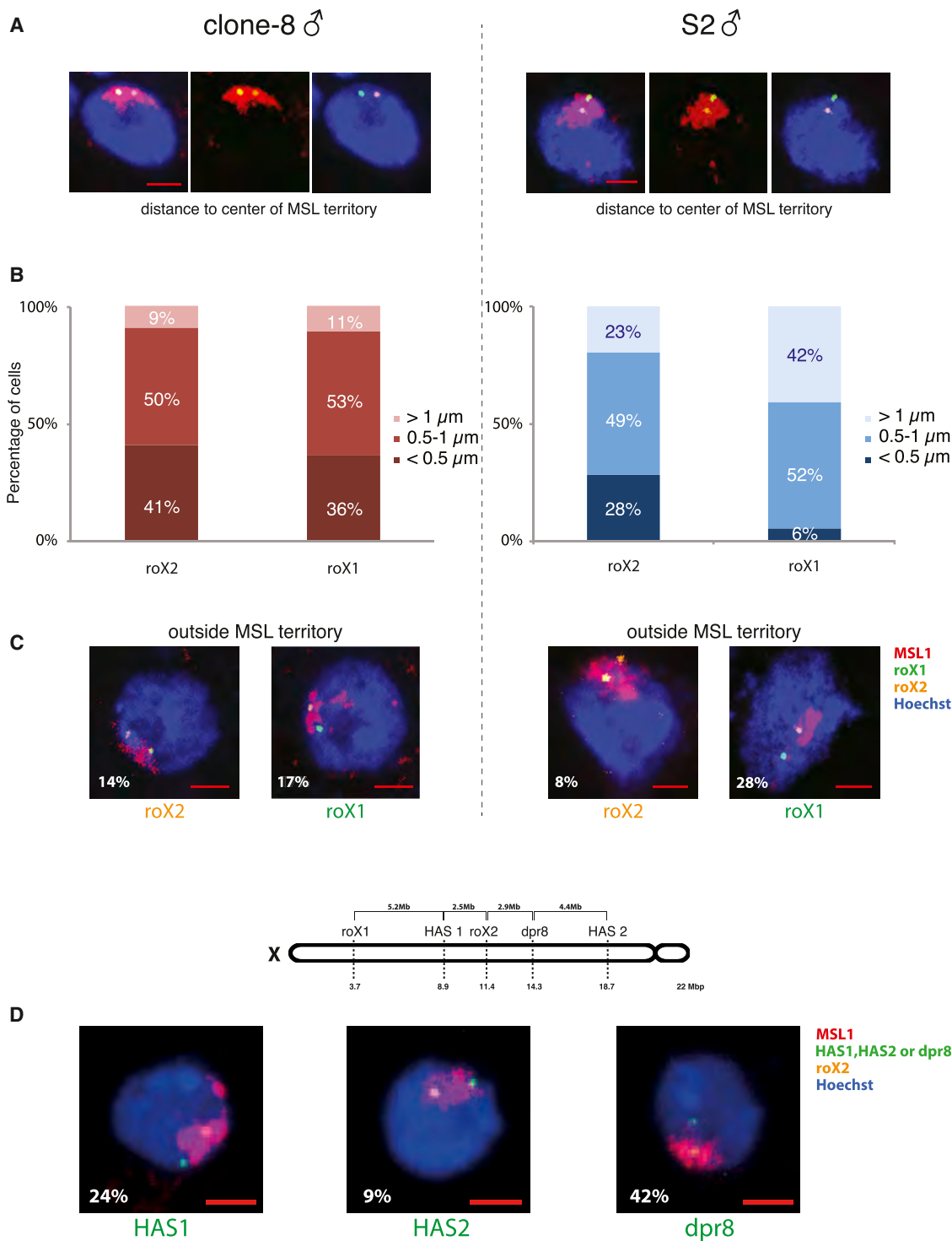
(B) Number of active genes in the vicinity (±15 kb) of HAS with or without long-range contacts.

(C) Gene expression (Cherbas et al., 2011) RNA-seq for genes close to roX1 and roX2. In clone-8 cells, roX1 and roX2 are active, whereas, in S2 cells, only roX2 is active.

(D) Schematic of long-range Hi-C contacts originating either from the roX1 (±25 kb) or the roX2 locus (±25 kb) in clone-8, S2, and Kc cells. The activity status for each of the roX genes is shown in the respective panels. Complementary images showing 4C contacts can be found in Figures S6E and S6F.

See also Figure S6 and Table S4.

A

clone-8 ♂

S2 ♂

distance to center of MSL territory

distance to center of MSL territory

B

Percentage of cells

100%

9%

50%

41%

roX2

11%

53%

36%

roX1

■ > 1 μm
■ 0.5-1 μm
■ < 0.5 μm

50%

0%

100%

23%

49%

28%

roX2

42%

52%

6%

roX1

■ > 1 μm
■ 0.5-1 μm
■ < 0.5 μm

50%

0%

C

outside MSL territory

outside MSL territory

MSL1
roX1
roX2
Hoechst

14%   roX2

17%   roX1

8%   roX2

28%   roX1

5.2Mb   2.5Mb   2.9Mb   4.4Mb

roX1   HAS 1   roX2   dpr8   HAS 2

X

3.7   8.9   11.4   14.3   18.7   22 Mbp

D

MSL1
HAS1,HAS2 or dpr8
roX2
Hoechst

24%   HAS1

9%   HAS2

42%   dpr8

Figure 4. Differential Positioning of HAS Relative to the MSL Territory

(A) Representative maximum intensity projections of confocal image stacks of 3D FISH experiments with *roX1* (green) and *roX2* (orange) probes in S2R+ and clone8 male *D. melanogaster* cells. MSL1 immunostaining is shown in red, and DNA is counterstained with Hoechst (blue). Scale bar, 2 μm.

(B) 3D distances from the center of mass of probes to the center of mass of the MSL territory. Graphs show the distribution of *roX1* and *roX2* probes in the different bins within the two cell lines: clone-8 (left) and S2R+ cells (right).

(C) Percentage of cells in which either *roX1* or *roX2* are localized outside the MSL territory, as demarcated by MSL1 immunostaining, in clone-8 (left) and S2R+ cells (right). Scale bars, 2 μm. n = 64 for clone-8 cells and 71 for S2R+ cells.

(legend continued on next page)

4C viewpoint on the translocated region, we observed long-range contacts with the X chromosome. The same viewpoint in Kc cells, lacking this translocation, did not show any contacts (Figure S7B). These data demonstrate that, by being physically associated to the X chromosome, the MSL complex can spread via long-range contact over a region lacking HAS.

The MSL Complex Can Target a HAS Independent of the Proximity of the roX RNA Production Site

As a testable prediction following the above observations, we expected that autosomes would also display MSL spreading when a *roX* gene is placed in the appropriate region. In an adaptation of a classical rescue experiment (Meller and Rattner, 2002), we next generated transgenic flies carrying as a sole source of roX lncRNA a *roX2* insertion on the right arm of the third chromosome (3R) at a precise position (86F8) in a *roX1/2* double mutant background. The inserted *roX2* gene was complemented with an array of lac operon sequences and an EGFP-fused lacI as a reporter to visualize the transgene without the need of DNA FISH (Supplemental Experimental Procedures). Consistent with previous reports (Kelley et al., 1999), roX RNA produced from the autosomal *roX* insertion properly targeted the X chromosome, enabling functional dosage compensation and restoring male viability. Moreover, the *roX* transgene was able to recruit the MSL complex to the ectopic insertion site, where, in addition to local spreading of the MSL complex into direct flanking regions on the autosome (seen in salivary gland polytene chromosomes), we repeatedly detected MSL binding at band 88B on the third chromosome (Figure 5B). Interestingly, this cytological position resides ~2.6 Mb away from the insertion site, with no detectable binding within an ~1.7-Mb region in between targeted bands, supporting an involvement of the 3D conformation. Although S2 and clone-8 cells are not expected to share the 3D structure of polytene chromosomes, it is quite remarkable that, out of many other possibilities, we detect an enrichment of Hi-C contacts in S2 (Figure S7C) and clone-8 cells (data not shown) between 86F and 88B. This suggests that HAS (such as *roX* loci) enable spreading of the MSL complex to interaction sites that can be far away on a linear scale and that do not necessarily need to be HAS themselves. We further confirmed this observation by two independent approaches involving the ectopic expression of *roX1* and *roX2* genes from an autosomal location (VK33 on chromosome 3L) in the roX1/2 double-null background: insertion of UAS-driven *roX1* or *roX2* genes and translocation of an X chromosomal segment containing either *roX1* or *roX2* genes to the ectopic site. In all cases, the autosomally expressed roX fully rescued male lethality and induced MSL targeting to the X chromosome, as seen in polytene chromosome spreads (Figure 5C). A non-functional roX RNA lacking important stem loop structures and expressed from VK33 under its native promoter, on the other hand, was not able to rescue male lethality and led to diminished or mis-targeting to the chromo-center (Ilik et al., 2013; Figure 5D; Figure S7D). These experiments verify that the properties of the ectopic *roX1/2* are independent of the promoter or location of the *roX* genes but require an intact roX structure.

It has been proposed that the X chromosomal territory forms in a self-organizing process around the *roX* genes by attracting HAS (Grimaud and Becker, 2010). However, such a model seems incompatible with the targeting of the X chromosome observed for polytene chromosomes in the previous experiments unless the autosomal *roX2* was looping into the X chromosomal territory to physically reach the X chromosome. To test this, we analyzed intact nuclei from the LacO-roX2 line where we observed the formation of a distinct additional MSL territory surrounding the ectopic insertion of *roX2*, frequently distant to the main MSL territory in three different tissues (salivary glands and fat body [examples of polytenic tissues] as well as imaginal discs [diploid tissue]) (Figure 5E). These results suggest that physical proximity is not necessary for the transfer of roX RNA from its production site to its targets on the X chromosome and that it can occur efficiently despite a possibly distinct 3D chromosome conformation of different tissues. Still, we also show that spreading to spatially proximal regions, situated at a long genetic distance, is possible because the extra MSL territory surrounding the insertion site appears discontinuous on polytene chromosome squashes (Figure 5B).

Ectopic Insertion of *roX*-HAS Leads to Transcriptional Upregulation of Neighboring Autosomal Genes In Vivo

To further investigate the spreading of the MSL complex from ectopic *roX2* HAS to nearby autosomal locations, we isolated RNA from flies with or without ectopic *roX2* HAS and measured the expression of ten genes upstream and ten genes downstream of the insertion site. Active genes surrounding the ectopic *roX2* HAS showed a consistent upregulation of expression in comparison with control flies (Figure 6A). To complement this analysis, we performed ChIP-seq with an antibody against MOF from the same transgenic larvae. On autosomes, MOF binds only to gene promoters, whereas, on the X chromosome, it also binds to gene bodies (Kind et al., 2008; Figure 6B, right). In transgenic flies carrying an ectopic *roX2*, MOF binds not only to the bodies of active genes surrounding the autosomal insertion point but also at a distance from it (Figure 6B, left and center). Importantly, we could detect MOF binding at the body of active genes up to 0.5 Mb upstream (Figure 6B) and 2 Mb downstream of the insertion site (data not shown). This suggests that the presence of a HAS on an ectopic autosomal location enables MOF to spread to proximal and distal regions on autosomes. To further validate the influence of ectopic *roX* insertion on the expression of neighboring genes, we also determined the expression of genes neighboring a different transgene used in this study (i.e., the X chromosomal translocation to 65B2 on chromosome arm 3L [VK33], where *roX2* is expressed under its native promoter)

(D) Three additional DNA FISH probes (*dpr8*, HAS1, and HAS2) were used in S2 cells, paired with roX2, to study their positioning with respect to the MSL territory. Representative pictures of each probe pair are shown, with the percentage of probes escaping the MSL territory indicated at the bottom left of each panel. Shown above is a schematic of the genomic location of all probes used in this study on the X chromosome of *D. melanogaster*. Scale bars, 2 μm. n = 45 for HAS1, for HAS2, and 30 for *dpr8*.
See also Table S5.
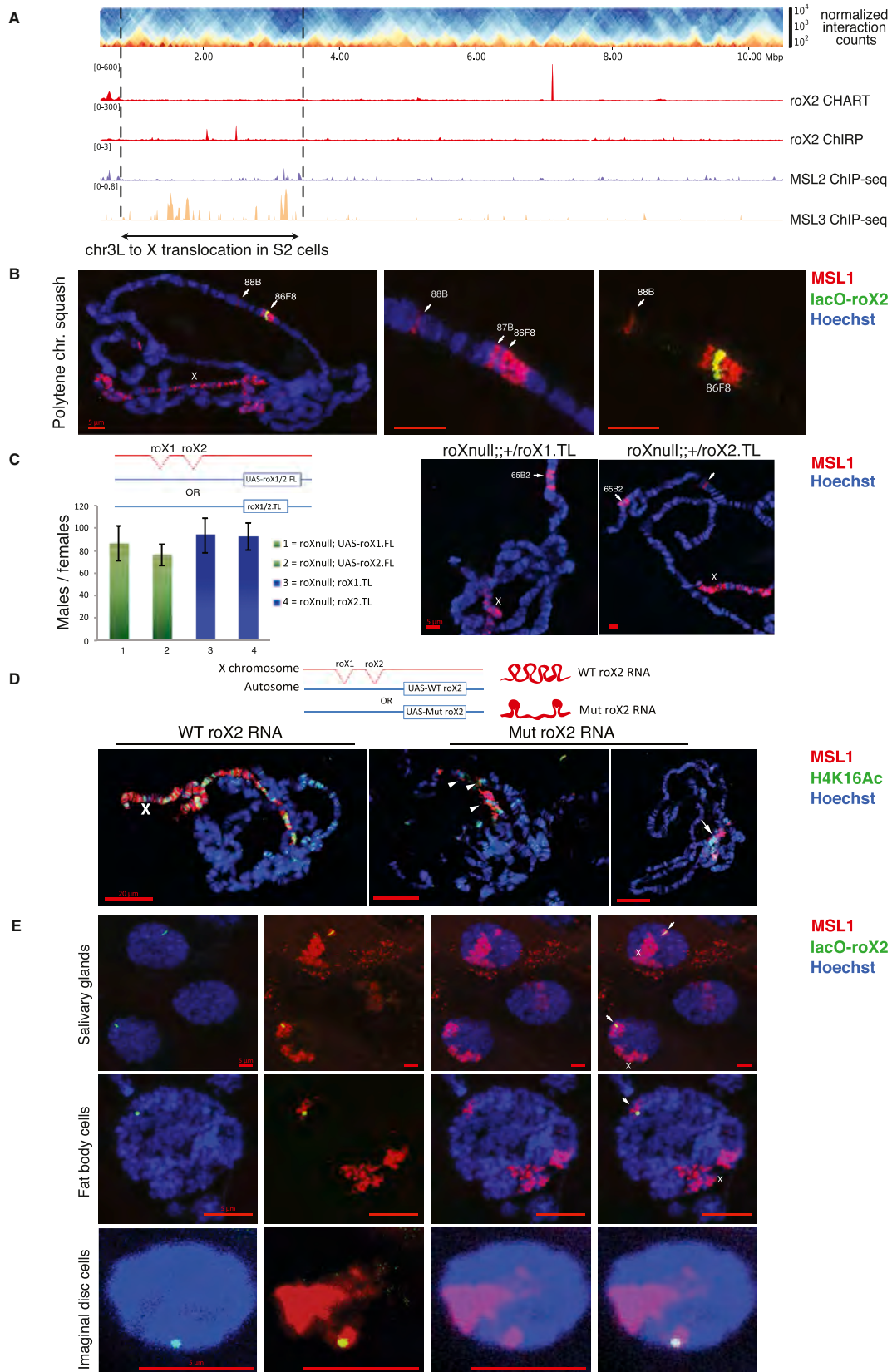
(legend on next page)

and also observed enhanced expression of ten genes surrounding this insertion site (Figure 6C). Taken together, these observations corroborate that the presence of a roX2 HAS enables the MSL complex to spread distally from the site of insertion and that it has activating potential on gene expression in vivo.

Nucleosome Positioning at HAS Is Dependent on the Presence of MSL2

Investigating functional HAS more closely, we observed that they tend to display increased DNaseI hypersensitivity (Figure S7E). We therefore asked whether the local chromatin structure of HAS can be influenced by the MSL complex. To this end, we mapped the positioning of nucleosomes in S2 cells by treating the chromatin with micrococcal nuclease, followed by deep sequencing (MNase-seq) (Mavrich et al., 2008). We found nucleosome-depleted regions on HAS, flanked by well positioned nucleosomes (Figure 7A). To test whether the MSL complex has any roles in maintaining the nucleosome configuration, we performed MNase-seq analysis in S2 cells depleted of MSL2. Strikingly, the nucleosome pattern around HAS was lost upon depletion of MSL2 (p < 2.2e-16; Supplemental Experimental Procedures), whereas it was preserved at the transcription start site (TSS). We conclude that the MSL complex is crucial for the maintenance of the local nucleosome arrangement, specifically at HAS.

DISCUSSION

This study provides a first step toward understanding the role of chromosome conformation in dosage compensation in *D. melanogaster.* We observe that HAS, the landing regions of the MSL complex on the X chromosome, frequently reside in proximity to TAD boundaries. We demonstrate that HAS are enriched in Hi-C contacts to each other and to other X chromosomal regions and that this organization remains comparable between male and female cells.

The Conformation-Based Affinity Model Explains MSL Complex Targeting and Spreading on the X Chromosome

Our analysis revealed that HAS are characterized by a combination of DNA sequence (MREs), chromatin state (active), and gene architecture, which drives the specificity of the MSL complex toward the X chromosome (Figure 7B, Targeting). Our data suggest that when the MSL complex binds to HAS, it then spreads (either via an active mechanism or via diffusion) to spatially close regions to place the histone H4 lysine 16 acetylation (H4K16ac) mark on active genes (Figure 7B, Spreading). We propose a "conformation-based affinity" model based on the strategic location of HAS at highly interconnected regions of the *D. melanogaster* X chromosome that efficiently distribute the MSL complex over the X chromosome by attracting the MSL complex to *cis*-interacting HAS on the X chromosome. This system ensures that only this chromosome is specifically and globally targeted. By spreading from those HAS over short (3D) distances, all active genes on the X chromosome are then reached and acetylated without influencing the autosomes. We suggest that this system is resilient to major perturbations, exemplified by the large autosomal insertion from chromosome 3L and the ectopic expression of the *roX* genes that produce viable cells and flies, respectively (Figure 5).

The MSL Complex Is Crucial for Maintaining Nucleosome-Free Regions at HAS

Our MNase-seq analysis shows, for the first time, a direct effect of the MSL complex on nucleosome organization specifically on HAS (Figure 7A) and not on the TSS, despite prominent binding of MSL1/2 to promoter regions. The MSL complex may act similar to a pioneer DNA binding protein (Magnani et al., 2011) to establish nucleosome patterns at HAS and may act on neighboring active regions rather than modifying TAD boundaries. This system may be unique to flies because the *Drosophila* dosage compensation evolved a fine-tuning transcription activation mechanism rather than a complete shutdown of gene transcription as seen in mammalian X chromosome inactivation. It would be very interesting to see how nucleosome positioning is affected upon Xist binding in mammals.

HAS Locate at Regions with Abundant Long-Range Contacts to Facilitate Spreading

Although many factors, including the CCCTC-binding factor (CTCF) as well as tRNA and housekeeping genes, have been shown to be enriched at boundaries (Dixon et al., 2012; Hou

---

Figure 5. Insertion of an Autosomal Fragment into the X Chromosome and of the *roX2* HAS into an Ectopic Autosomal Location

(A) Hi-C heatmap of chromosome 3L highlighting a 2.67-Mb translocation (chr3L, ∼796,745–3,468,912) into chromosome X. Figure S7B shows 4C long-range contacts from the translocation to chromosome X.

(B) The autosomal insertion of LacO-roX2 (green) into 86F8 on chromosome arm 3R efficiently targets the MSL complex (MSL1, red) to the X chromosome in males and causes local spreading from the insertion site in polytene chromosomal immunostaining of salivary glands from male third-instar larvae. DNA is stained with Hoechst 33342 (blue). Figure S7C contains the Hi-C counts from the insertion point on band 86F8 in S2 cells. Scale bars, 5 μm.

(C) Left: rescue of roX1[SMC17A], roX2[Δ] double mutant (roX[null]) male-specific lethality by daGal4-induced expression from a full-length (FL) UAS-roX1 transgene (#1) or UAS-roX2 transgene (#2) or by expression driven by endogenous promoters upon translocation (TL) of roX1 (#3) or roX2 (#4) to the third chromosome. All transgenes are inserted in the same autosomal location (VK33). Transgenic males surviving to adulthood were counted and normalized to female siblings (100% viable, ∼1,100 flies/genotype). Columns show averages of n ≥ 3 separate crosses ± SD. See Experimental Procedures for details of the genetic crosses. Below the graph is a schematic of the transgenes used. Right: two representative polytene chromosome squashes are shown for roX[null]; TLroX1/2. Scale bars, 5 μm.

(D) A mutant form of roX RNA lacking important stem-loop structures is expressed from the VK33 autosomal insertion. This causes inefficient targeting of the MSL complex (MSL1, red) to the X chromosome and mislocalization of H4K16 acetylation (green), as visualized on polytene chromosomal immunostaining of salivary glands from male third-instar larvae (see Supplemental Experimental Procedures and Figure S7D for further details). Scale bars, 20 μm.

(E) As (C) but showing intact nuclei of whole salivary glands (top), fat body cells (center), and imaginal disc diploid cells (bottom). X indicates the X chromosome, and arrowheads point to the region around the lacO-roX2 autosomal insertion. Scale bars, 5 μm.

See also Figure S7.
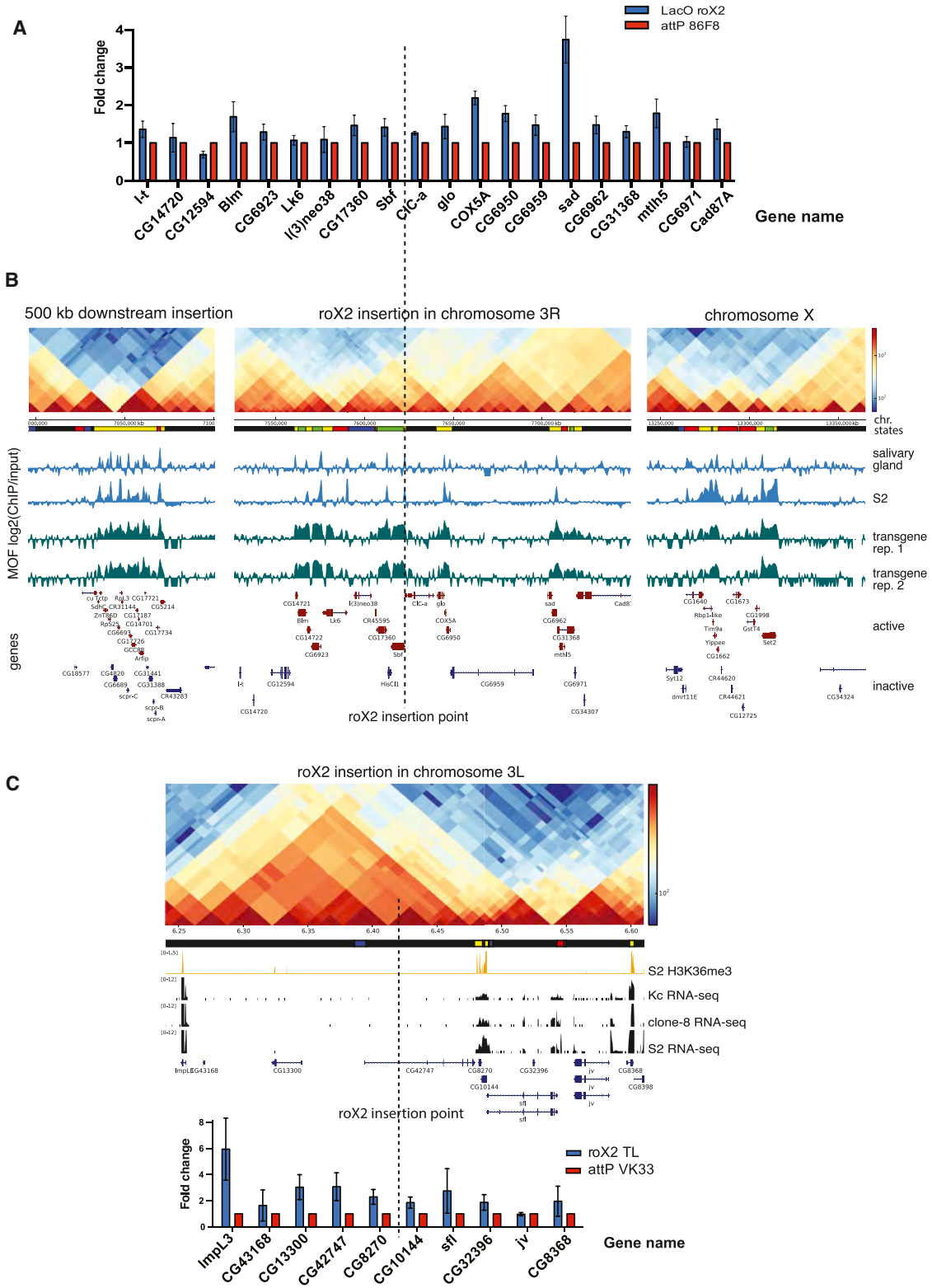
Figure 6. Ectopic Insertion of a HAS Leads to Spreading of MOF to and Enhanced Expression of Neighboring Autosomal Genes

(A) A HAS on a autosomal location causes an upregulation of autosomal genes. Expression of autosomal genes neighboring *roX2* insertion was analyzed by qRT-PCR. Flies analyzed were *roX* double mutant (*roX*[null]) combined with either a transgene carrying LacO-roX2 (under the endogenous promoter, LacO-roX2, blue

(legend continued on next page)

et al., 2012; Sexton et al., 2012), by dissecting the targeting and spreading activity of the MSL complex for the X chromosome we offer a plausible explanation behind the advantages of HAS localization. HAS are enriched at the X chromosomal boundaries and not at autosomal boundaries, where all other boundary factors will bind indiscriminately. Furthermore, we found that the few HAS that are not near a boundary also occupy locations of an elevated number of long-range contacts (Figure S5A), indicating that HAS form interaction hubs for the spreading of the MSL complex.

Activity of *roX* Genes Is Associated with a Higher Abundance of Long-Range Contacts

Hi-C as well as in vivo immunofluorescence show that active *roX* genes have more contacts and are closer to each other than inactive regions (Figure 3). These observations are in line with previous reports showing that active chromatin compartments interact more often with each other (Lieberman-Aiden et al., 2009; Sexton et al., 2012) and that active chromatin localizes to the interface of the chromosomal territory (Nagano et al., 2013). Our results imply that different transcriptional programs in each cell line or tissue are likely to be associated with a particular arrangement of long-range contacts, suggesting that the dosage compensation must be flexible to act over such diverse conformations without disturbing them. This idea is consistent with the observation that the chromosome conformation remains unchanged after knockdown of the MSL complex (Figures S6A–S6D), and stays in contrast to mammalian X inactivation, which involves chromatin condensation, gene inactivation, and alterations in chromosome conformation (Nora et al., 2012).

lncRNAs Work Differently in Coordinating Fly and Mammalian Dosage Compensation

Dosage compensation mechanisms in flies and mammals lead to opposite outcomes; namely, gene activation versus gene repression. However, both systems use lncRNAs transcribed from the dosage-compensated X chromosome. roX1 and roX2 RNA are expressed from the male hyperactivated X chromosome in *D. melanogaster*, whereas Xist is expressed from the inactivated X chromosome in mammalian females (Brockdorff and Turner, 2015; Grimaud and Becker, 2010). Recent work has shown that Xist spreads to distal sites on the X chromosome. Interestingly, this spreading is dependent on the spatial proximity of sites distal to the *Xist* gene (Engreitz et al., 2013; Simon et al., 2013). This is further exemplified by ectopic expression of

Xist from chromosome 21, where Xist spread only in *cis* on this chromosome (Jiang et al., 2013). In our study, ectopic insertion of *roX* transgenes on autosomes demonstrated that the roX/MSL complex can reach the X chromosome and rescue male lethality (Figures 5B–5D). Therefore, acting in *trans* is a special feature of roX RNAs (in conjunction with the MSL complex) not observed for Xist, indicating that the two systems utilize the respective lncRNAs differently. In both systems, however, the lncRNAs need to be functional because the stem loop structures of the roX RNAs are required for dosage compensation in *D. melanogaster* (Ilik et al., 2013; Figure S7D), whereas Xist needs the "A repeat domain" to induce mammalian X chromosome inactivation (Engreitz et al., 2013). The distinct mechanisms utilized by the Xist and roX RNAs exemplify the great versatility by which lncRNAs can be involved in the global regulation of single chromosomes and might reflect important differences between the two systems. In mammals, only one of the two X chromosomes needs to be inactivated. Therefore, a *trans* action of Xist RNA on the sister X chromosome would be detrimental to the organism. In contrast, the dosage-compensated X chromosome is present singularly in males in *Drosophila*. However, because the roX RNAs can act in *trans*, it may be disadvantageous to target the activating MSL complex to active genes on autosomes, hence the need for specific target regions (the HAS) unique to the X chromosome.
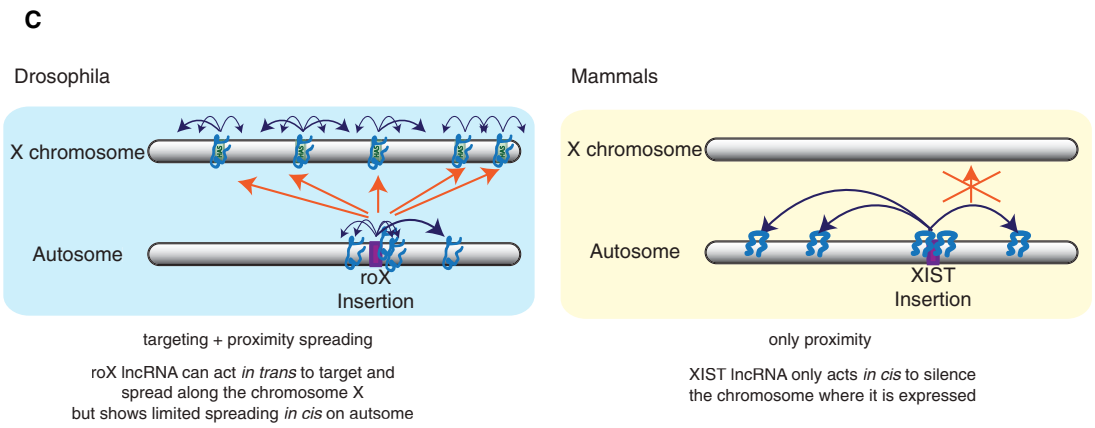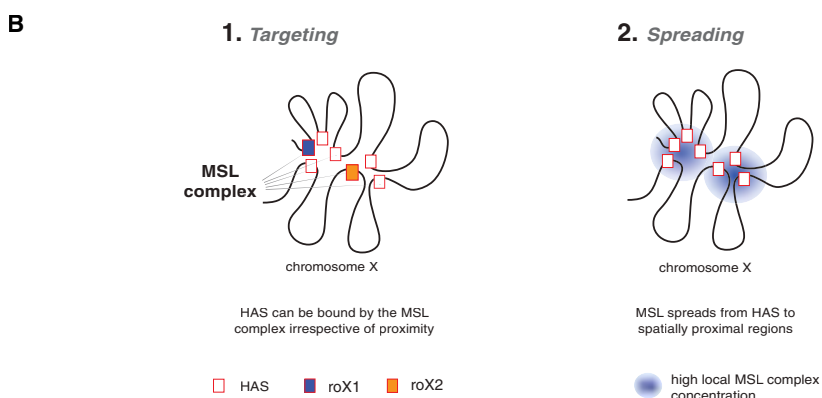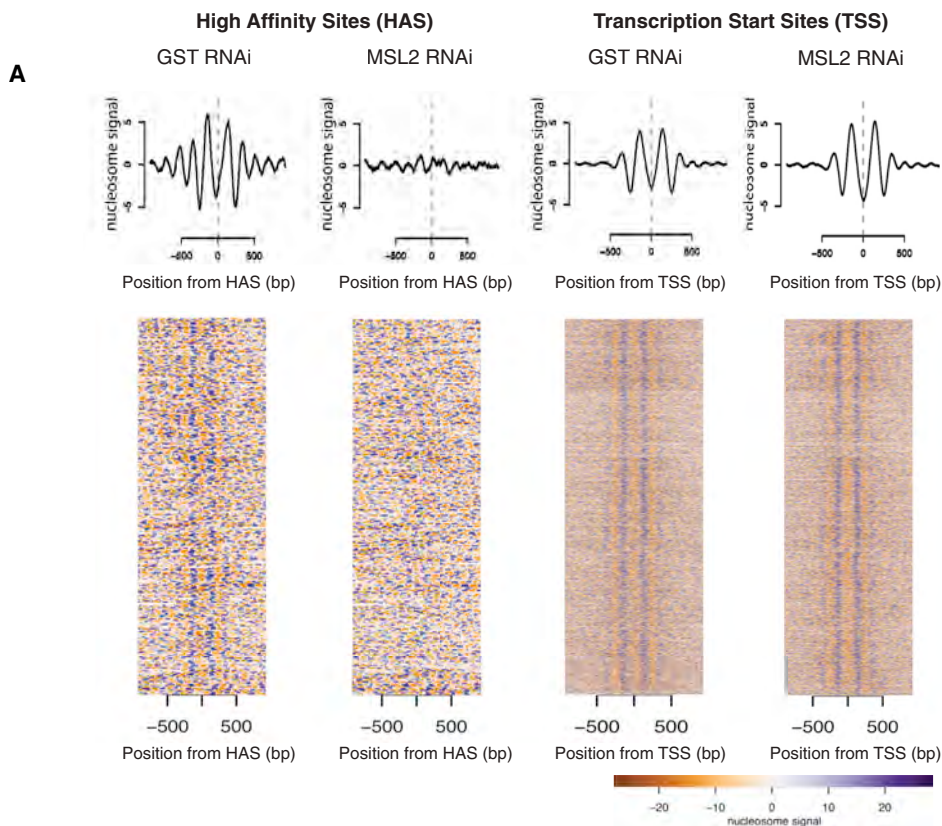
Faster Evolution of the X Chromosome May Favor Positioning of HAS to Interaction Hubs

To fully understand the occurrence of HAS at sites with extensive long-range interactions on the X chromosomes, it could be helpful to consider evolutionary models proposing that X chromosomes tend to evolve faster than autosomes (faster X effect) (Vicoso and Charlesworth, 2006). Under the faster X effect, traits only beneficial for males can introduce significant changes specific to the X chromosome on a short evolutionary timescale (Parsch and Ellegren, 2013; Vicoso and Charlesworth, 2006). Based on these and other observations suggesting that the X chromosome in flies is different from autosomes (Alekseyenko et al., 2012; Gallach, 2014; Meisel and Connallon, 2013; Zhang and Oliver, 2010), we assume that selective pressures on males favored the occurrence of HAS at regions of increased interactions, like TAD boundaries. Future analyses of different *Drosophila* species will open exciting opportunities to study the evolutionary changes of HAS in the context of X chromosomal architecture. Moreover, conformation-based affinity could be a

bars) or the attP transposon only in the same insertion site (AttP 86Fb, red bars). Relative gene expression levels of LacO-roX2 flies were calculated over that of AttP 86Fb flies. Data are represented as the mean of triplicates ± SEM. PFK levels were used as control.

(B) MOF spreads to distal regions from the *roX* insertion site on chromosome 3R. The tracks shown from top to bottom are as follows: Hi-C contacts from S2; chromatin states (Filion et al., 2010); input-normalized MOF ChIP-seq signals for third-instar larva salivary gland (Conrad and Akhtar, 2011); and S2 cells (Straub et al., 2013) and LacO-roX2 flies (two replicates). The Hi-C heatmap contains corrected counts from single restriction fragments. In the chromatin states track, the yellow and red regions represent active chromatin, black corresponds to repressive chromatin, and blue and green represent heterochromatin. The tracks are shown for a 100-kb region located 500 kb downstream of the *roX2* insertion site (left), a 250-kb region around the insertion site (center), and a 100-kb region on the X chromosome (right). The dotted line depicts the autosomal insertion site of *roX2* (chromosome 3R, 86F8).

(C) Expression of autosomal genes neighboring the *roX2* insertion was analyzed by qRT-PCR. Flies analyzed were *roX* double mutant (*roX*null) combined with either a translocation of *roX2* (roX 2 TL) to an insertion site on chromosome 3L (VK33) or the attP transposon alone in the same VK33 insertion site. Relative gene expression levels (± SEM) of roX2 TL flies were calculated over that of AttP VK33 flies. Data are represented as the mean of triplicates ± SEM. PFK levels were used as control. Genes in repressive chromatin or heterochromatin have low or no upregulation (e.g., *CG14720* and *CG15594* in A and *CG43168* and jv in C). H3K36me3 and RNA-seq tracks are used to indicate gene transcription activities in different cell lines.

A

High Affinity Sites (HAS)

GST RNAi   MSL2 RNAi

Transcription Start Sites (TSS)

GST RNAi   MSL2 RNAi

Position from HAS (bp)   Position from HAS (bp)   Position from TSS (bp)   Position from TSS (bp)

B

1. *Targeting*

MSL complex

chromosome X

HAS can be bound by the MSL complex irrespective of proximity

☐ HAS   ■ roX1   ■ roX2

2. *Spreading*

chromosome X

MSL spreads from HAS to spatially proximal regions

● high local MSL complex concentration

C

Drosophila

X chromosome

Autosome

roX Insertion

targeting + proximity spreading

roX lncRNA can act *in trans* to target and spread along the chromosome X but shows limited spreading *in cis* on autsome

Mammals

X chromosome

Autosome

XIST Insertion

only proximity

XIST lncRNA only acts *in cis* to silence the chromosome where it is expressed

(legend on next page)

generic mechanism for other regulatory elements to exert their functions. It remains to be seen in which contexts the in *cis* versus in *trans* action of different lncRNAs is essential for their function and how chromosome conformation, long-range contacts, HAS, and regulation of transcription have co-evolved for dosage compensation.

EXPERIMENTAL PROCEDURES

Hi-C Experimental Procedure and Analysis

Hi-C in S2 or clone-8 cells using HindIII as restriction enzyme was carried out as described by Belton et al. (2012) with the following minor modifications. Starting material for all samples was 50 million insect cells/sample. After lysis, samples were taken up in 125 μl and split into two aliquots of 50 μl (Hi-C samples), and the remaining 25 μl (3C control) were used to adjust for the smaller size of the *Drosophila* genome compared with mammalian cells. Accordingly, for each 3C control, only half of the volumes per tube were used compared with the original protocol. For details, see Supplemental Experimental Procedures.

4C-Seq Experimental Procedure and Analysis

4C-seq in S2 and Kc cells was carried out as described by Splinter et al. (2012) with minor modifications as follows. 50–100 million S2 cells or Kc cells, fixed as described above, were used for two biological replicates. DpnII (New England Biolabs) was used as the primary and Csp6I (Thermo Scientific) as the secondary restriction enzyme. For each viewpoint, two 160-ng PCR reactions (8 cycles with a 55°C annealing temperature, followed by 18 cycles at 63°C) were prepared and cleaned up. All different viewpoint libraries for one biological replicate were mixed equimolarly and sequenced on separate lanes on an Illumina HiSeq2500 DNA sequencer. Primer sequences and coordinates for the experiments can be found in Table S4.

Fly Culture and Genetics

Details of fly culture conditions and genetics are explained in the Supplemental Experimental Procedures.

FISH

Fluorescence in situ hybridization procedures were performed as described previously (Vaquerizas et al., 2010). To perform DNA FISH, approximately ten 5-kb regions were chosen in the genome and amplified by PCR from genomic DNA with five to ten primers pairs, each covering around 0.5–3 kb. Primer sequences are available upon request. The roX1 probe sequence was taken from Vaquerizas et al. (2010). The *roX2* and *dpr8* probe sequences were taken from Grimaud and Becker (2009). The HAS1 and HAS2 probes were designed and generated for this study.

Nucleosome Positioning Analysis in *Drosophila*

Nucleosome positioning analysis using MNase-seq was performed essentially as described in Mavrich et al. (2008). For details of the analysis, see the Supplemental Experimental Procedures.

REFERENCES

Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.-K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J., et al. (2008). A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. Cell *134*, 599–609.

Alekseyenko, A.A., Ho, J.W.K., Peng, S., Gelbart, M., Tolstorukov, M.Y., Plachetka, A., Kharchenko, P.V., Jung, Y.L., Gorchakov, A.A., Larschan, E., et al. (2012). Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context. PLoS Genet. *8*, e1002646.

Belton, J.-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods *58*, 268–276.

Figure 7. Depletion of the MSL Complex Severely Affects Nucleosome Positioning at HAS but not at the TSS

(A) Summary plots (top) and heatmaps (bottom) showing normalized nucleosome occupancy for regions centered around HAS (left, n = 257) and active TSSs (right, n = 5985) in control (GST RNAi) and MSL2-depleted cells (MSL2 RNAi). A bar representing the nucleosome signal is shown below.

(B) Conformation-based affinity model. We propose a model in which chromosome X is targeted via HAS by the MSL complex independently of spatial proximity (1). Then the complex spreads from the HAS to spatially proximal regions (2). In (1), the MSL complex binds specific regions of chromosome X containing a sequence motif (the MRE) if this appears at the end of active genes that are at TAD boundaries. In (2) the complex spreads (probably by diffusion) from HAS to spatially close regions. Because TAD boundaries appear enriched in contacts within each other as well as to other regions, it is expected that they are physically close, therefore offering an optimal location from which to reach all active genes on chromosome X.

(C) Comparison of the ectopic expression of roX RNA versus Xist RNA on autosomes. Blue arrows denote spreading of the lncRNA in 3D proximity, and orange arrows denote targeting of the lncRNA in *trans* (blue box). In *Drosophila*, ectopically expressed roX displays restricted spreading on the autosome, indicating that spatial proximity is beneficial, but not sufficient, for roX RNA spreading. Importantly, roX RNA expressed from an autosomal location specifically targets the X chromosome in *trans*, leading to coating of the entire X chromosome by the roX/MSL complex independent of spatial proximity (yellow box). In mammals, ectopically expressed Xist RNA can spread and coat the autosome using spatial proximity (Engreitz et al., 2013; Jiang et al., 2013; Simon et al., 2013) without targeting the X chromosome in *trans*.

See also Figure S7.

Brockdorff, N., and Turner, B.M. (2015). Dosage compensation in mammals. Cold Spring Harb. Perspect. Biol. *7*, a019406.

Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., Dumais, J., et al. (2011). The transcriptional diversity of 25 Drosophila cell lines. Genome Res. *21*, 301–314.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell *44*, 667–678.

Conrad, T., and Akhtar, A. (2011). Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat. Rev. Genet. *13*, 123–134.

Currie, D.A., Milner, M.J., and Evans, C.W. (1988). The growth and differentiation in vitro of leg and wing imaginal disc cells from Drosophila melanogaster. Development *102*, 805–814.

de Wit, E., Bouwman, B.A.M., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J.A.M., Krijger, P.H.L., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227–231.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212–224.

Gallach, M. (2014). Recurrent turnover of chromosome-specific satellites in Drosophila. Genome Biol. Evol. *6*, 1279–1286.

Gilfillan, G.D., Straub, T., de Wit, E., Greil, F., Lamm, R., van Steensel, B., and Becker, P.B. (2006). Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex. Genes Dev. *20*, 858–870.

Grimaud, C., and Becker, P.B. (2009). The dosage compensation complex shapes the conformation of the X chromosome in Drosophila. Genes Dev. *23*, 2490–2495.

Grimaud, C., and Becker, P.B. (2010). Form and function of dosage-compensated chromosomes–a chicken-and-egg relationship. BioEssays *32*, 709–717.

Heard, E., and Disteche, C.M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Genes Dev. *20*, 1848–1867.

Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol. Cell *48*, 471–484.

Ilik, I.A., Quinn, J.J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R.C., Luscombe, N., Backofen, R., et al. (2013). Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila. Mol. Cell *51*, 156–173.

Jiang, J., Jing, Y., Cost, G.J., Chiang, J.-C., Kolpa, H.J., Cotton, A.M., Carone, D.M., Carone, B.R., Shivak, D.A., Guschin, D.Y., et al. (2013). Translating dosage compensation to trisomy 21. Nature *500*, 296–300.

Johansson, A.-M., Allgardsson, A., Stenberg, P., and Larsson, J. (2011). msl2 mRNA is bound by free nuclear MSL complex in Drosophila melanogaster. Nucleic Acids Res. *39*, 6428–6439.

Keller, C.I., and Akhtar, A. (2015). The MSL complex: juggling RNA-protein interactions for dosage compensation and beyond. Curr. Opin. Genet. Dev. *31*, 1–11.

Kelley, R.L., Meller, V.H., Gordadze, P.R., Roman, G., Davis, R.L., and Kuroda, M.I. (1999). Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell *98*, 513–522.

Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P., and Akhtar, A. (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. Cell *133*, 813–828.

Liang, J., Lacroix, L., Gamot, A., Cuddapah, S., Queille, S., Lhoumaud, P., Lepetit, P., Martin, P.G.P., Vogelmann, J., Court, F., et al. (2014). Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. Mol. Cell *53*, 672–681.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Magnani, L., Eeckhoute, J., and Lupien, M. (2011). Pioneer factors: directing transcriptional regulators within the chromatin environment. Trends Genet. *27*, 465–474.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.

Meisel, R.P., and Connallon, T. (2013). The faster-X effect: integrating theory and data. Trends Genet. *29*, 537–544.

Meller, V.H., and Rattner, B.P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J. *21*, 1084–1091.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59–64.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Parsch, J., and Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene expression. Nat. Rev. Genet. *14*, 83–87.

Quinn, J.J., Ilik, I.A., Qu, K., Georgiev, P., Chu, C., Akhtar, A., and Chang, H.Y. (2014). Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat. Biotechnol. *32*, 933–940.

Quinodoz, S., and Guttman, M. (2014). Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. Trends Cell Biol. *24*, 651–663.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Schneider, I. (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J. Embryol. Exp. Morphol. *27*, 353–365.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. USA *108*, 20497–20502.

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature *504*, 465–469.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J. *32*, 3119–3129.

Splinter, E., de Wit, E., van de Werken, H.J.G., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods *58*, 221–230.

Straub, T., Grimaud, C., Gilfillan, G.D., Mitterweger, A., and Becker, P.B. (2008). The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. PLoS Genet. *4*, e1000302.

Straub, T., Zabel, A., Gilfillan, G.D., Feller, C., and Becker, P.B. (2013). Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq. Genome Res. *23*, 473–485.

Thomas-Chollier, M., Hufton, A., Heinig, M., O'Keeffe, S., Masri, N.E., Roider, H.G., Manke, T., and Vingron, M. (2011). Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nat. Protoc. *6*, 1860–1869.

Van Bortle, K., Nichols, M.H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol. *15*, R82.

Vaquerizas, J.M., Suyama, R., Kind, J., Miura, K., Luscombe, N.M., and Akhtar, A. (2010). Nuclear pore proteins nup153 and megator define transcriptionally active regions in the Drosophila genome. PLoS Genet. *6*, e1000846.

Vicoso, B., and Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. *7*, 645–653.

Zhang, Y., and Oliver, B. (2010). An evolutionary consequence of dosage compensation on Drosophila melanogaster female X-chromatin structure? BMC Genomics *11*, 6.

Zhang, Y., Malone, J.H., Powell, S.K., Periwal, V., Spana, E., Macalpine, D.M., and Oliver, B. (2010). Expression in aneuploid Drosophila S2 cells. PLoS Biol. *8*, e1000320.

Supplemental Information

High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex

across the X Chromosome in *Drosophila*

Fidel Ramírez, Thomas Lingg, Sarah Toscano, Kin Chung Lam, Plamen Georgiev, Ho-Ryun Chung, Bryan R. Lajoie, Elzo de Wit, Ye Zhan, Wouter de Laat, Job Dekker, Thomas Manke, Asifa Akhtar
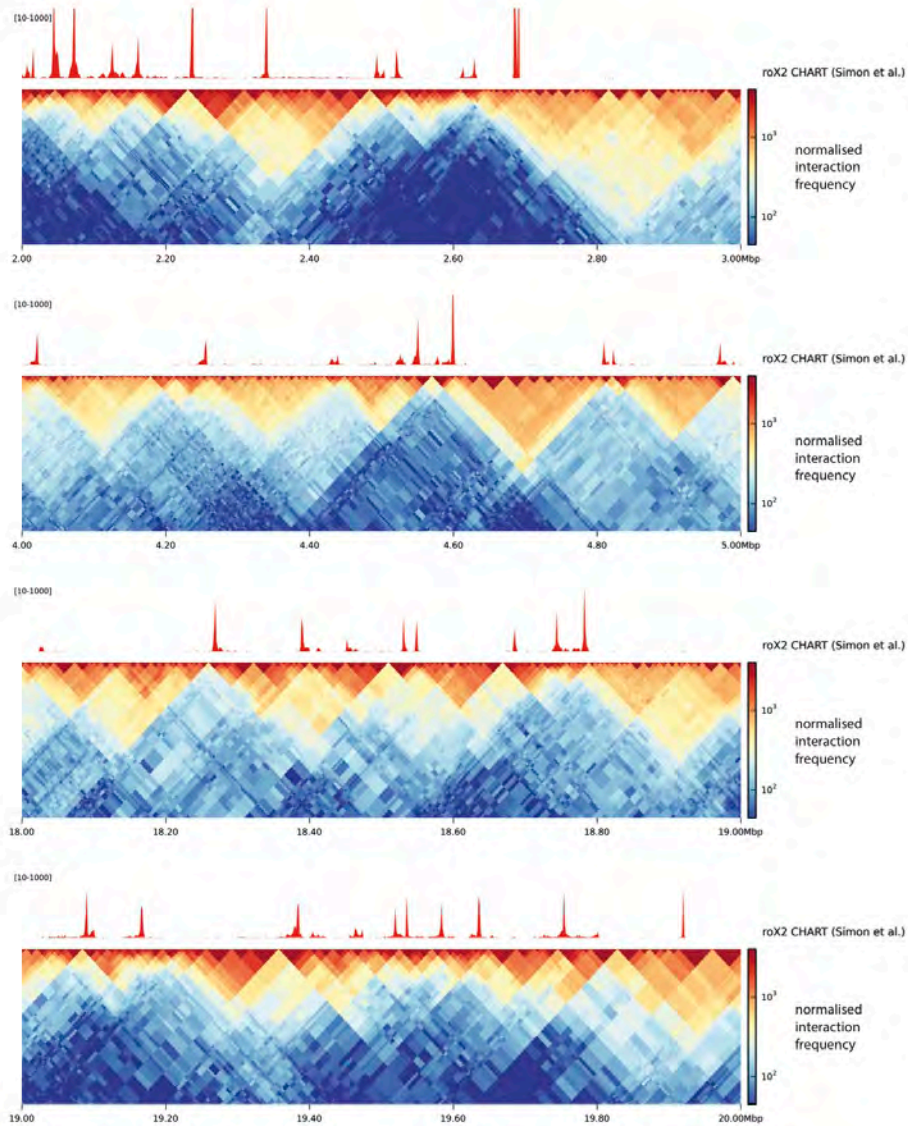
Figure S1
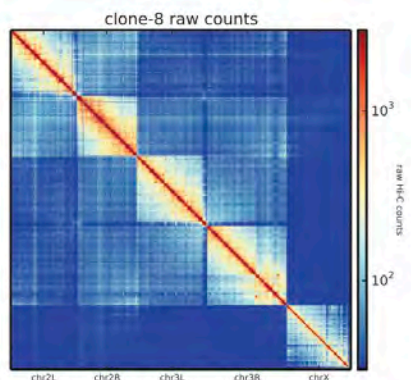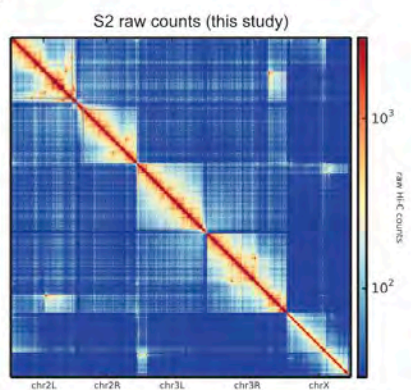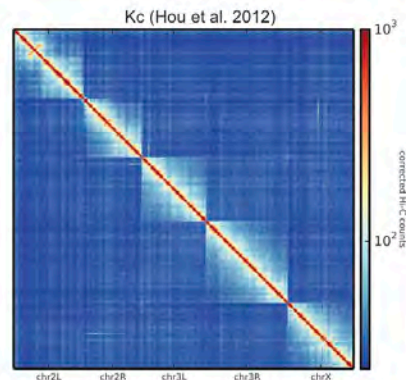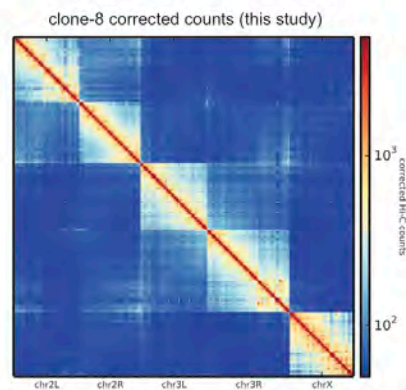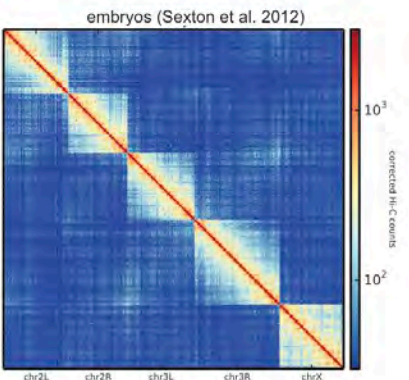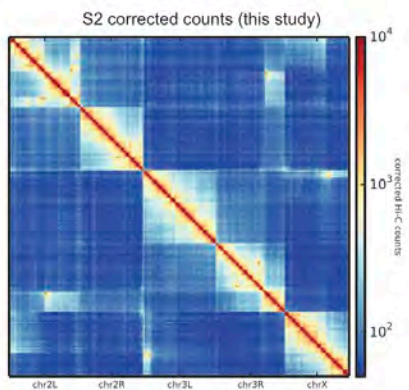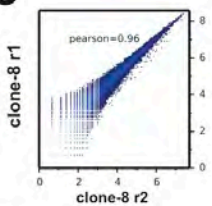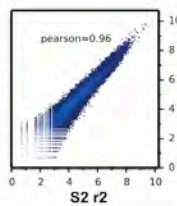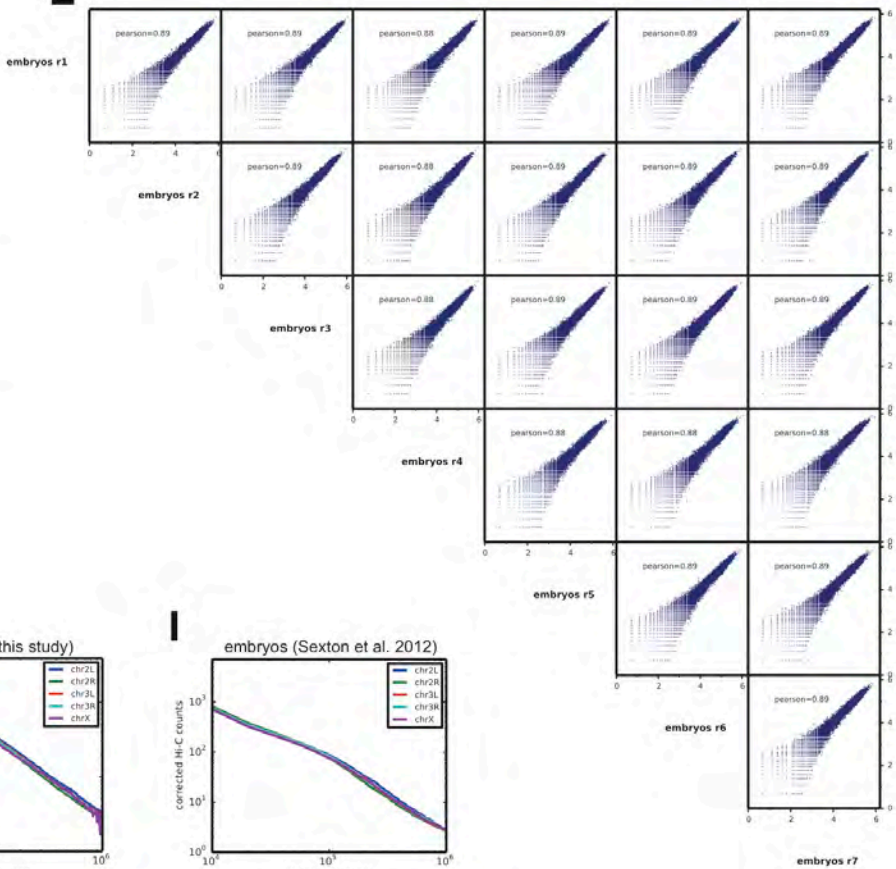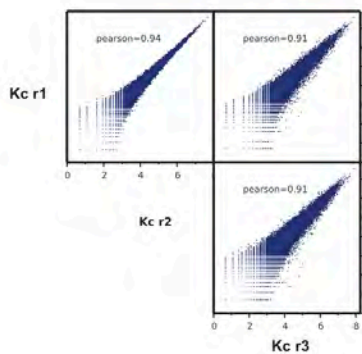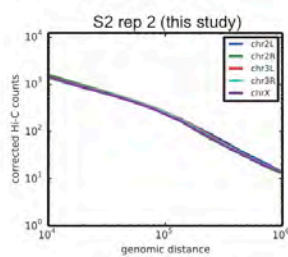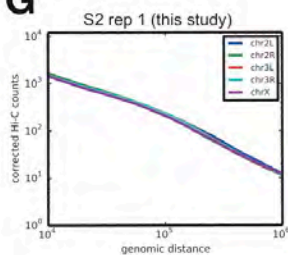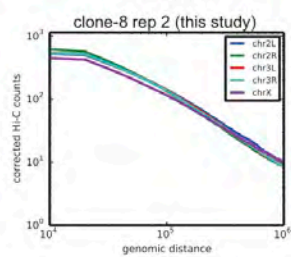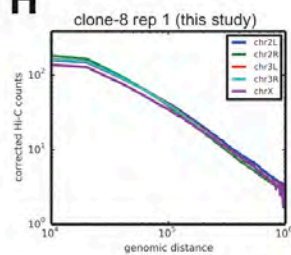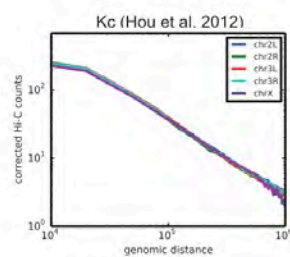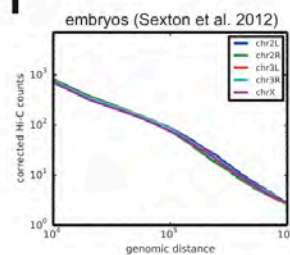
Figure S1.

Validation of corrected Hi-C data, related to Figure 1.

A. Hi-C Matrices showing uncorrected Hi-C counts for S2 and clone-8 cell lines. **B.** Corrected matrices using the method from (Imakaev et al., 2012). In this panel we also show the Hi-C matrices for Kc cells (Hou et al., 2012) and *Drosophila* embryos (Sexton et al., 2012) that were processed using the same pipeline as the S2 and clone-8 data. All matrices depicted are based on 100 kb binning. **C-F.** The correlations were based on Hi-C counts using 10 kb bins. Only counts between bins smaller than 300 kb were studied; regions beyond this range are noisy at this resolution. All cases in which both of the matrices had zero counts were excluded. **C**. Clone-8 correlation, **D**. S2 correlation, **E**. embryos correlation, **F**. Kc correlation. Correlation of corrected Hi-C counts resulted in similar values for all cases. **G-I.** Power law decay of Hi-C counts with genomic distance. For all plots 10 kb bins were used. **G**. plots for Hi-C data from each of the two S2 cell lines replicates. **H.** plots for Hi-C data from each of the two clone-8 replicates. **I**. Plots for Hi-C data for the merge of embryos (upper) and Kc cells (lower) replicates.

Figure S2

A

B
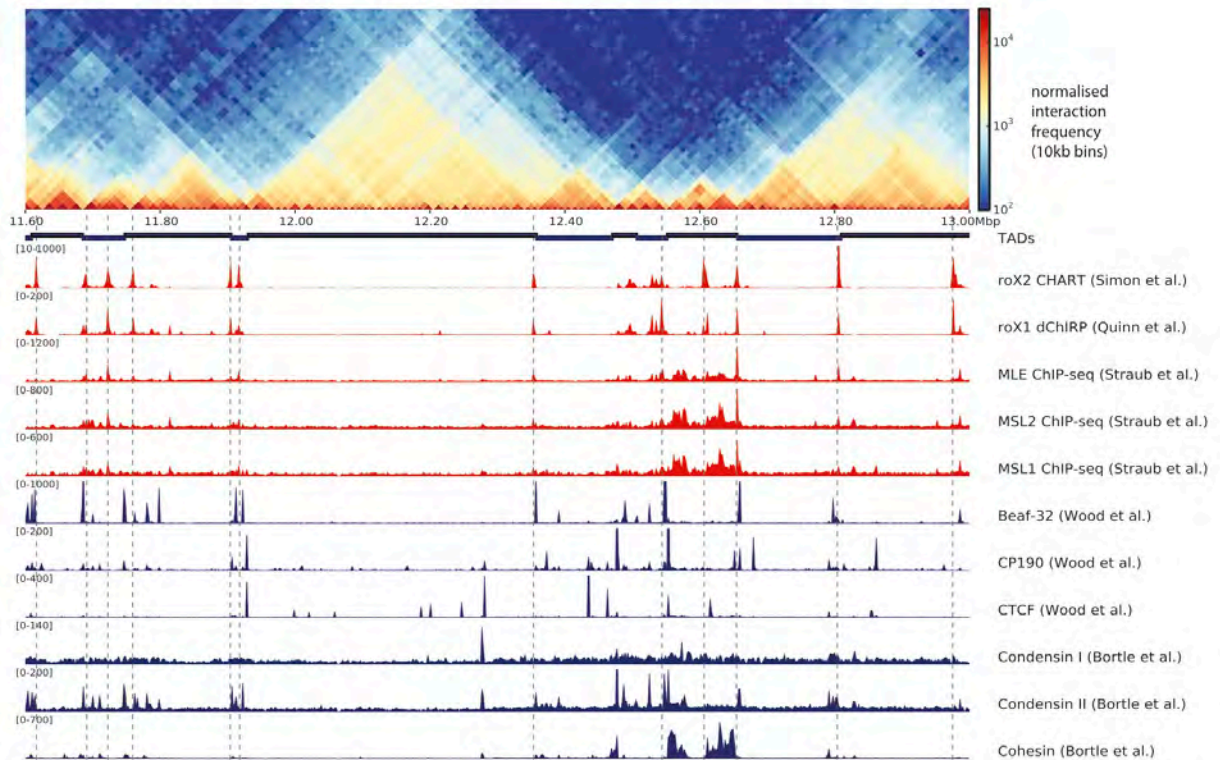
Figure S2.

Analysis of TAD boundaries, related to Figure 1.

A. TAD boundaries and roX2. roX2 CHART data (Simon et al., 2011) and Hi-C data from S2 cells at restriction fragment resolution (each bin about 3.500 bp). The different panels show the roX2 signal correlating with boundaries at four different random positions of 1 Mb along the X chromosome. **B.** TAD boundaries and insulator proteins. Similar to Fig. 1A but additionally the dosage compensation complex members (red tracks), insulator (Wood et al., 2011) and architectural proteins (Van Bortle et al., 2014) are shown (blue tracks).

Figure S3

A TAD score = 1 / min(1+2, 1+3)

B TAD separation score — chromosome 3L

C
- embryos TAD sep. score
- APBS
- embryos
- S2
- clone-8
- Kc
- Sexton et al.
- Hou et al.
- Chromatin states

D
- embryos
- kc
- s2
- clone-8

TAD separation score
- embryos
- kc
- s2
- clone-8
- APBS
- Chromatin states

Chromosome 3R

E

F

G
Sexton et al.
distance from HAS to boundary

Hou et al.
distance from HAS to boundary

H Embryos   Kc   S2   clone-8

I H3K4me2   H3K79me3   TBP   chromatin states
- between inactive
- active to inactive
- inactive to active

Figure S3.

TAD separation score, related to Figure 1.

A. Illustration for the computation of the TAD separation score. The score, computed at the position where the triangles (2) and (3) meet, is the ratio of the total corrected Hi-C contacts in (1) to the minimum of the total Hi-C contacts from (2) and (3). The score aims to represent the inter-TAD counts normalized by intra-TAD counts. **B.** The TAD separation score reaches a local minimum between TADs when computed along the entire chromosome. The score also informs about the degree of TAD separation: adjacent TADs with more of Hi-C contacts between them receive a larger score compared to TADs having fewer contacts. **C.** Comparison of different methods to identify TADs. On top we show the embryos TAD separation score, next the architectural protein binding sites (Van Bortle et al., 2014) that are thought to define boundaries, next are the positions of the boundaries identified using the TAD separation score in embryos, S2, clone-8 and Kc cells. Following are the boundary positions originally reported for embryos (Sexton et al., 2012) and for Kc cells (Hou et al., 2012). The last track corresponds to the chromatin states as defined in (Filion et al., 2010). The vertical blocks highlight the positions for which there is a good agreement between the different methods and the different cell types. **D.** Comparison of TADs from the different cell types and embryos. TAD boundaries tend to be very conserved within cell types as shown by the corrected Hi-C counts and by the TAD separation score. The vertical lines highlight the agreement between the different TADS. The red vertical line close to the center shows a case in which a weak (i.e. intra-TAD contacts are ele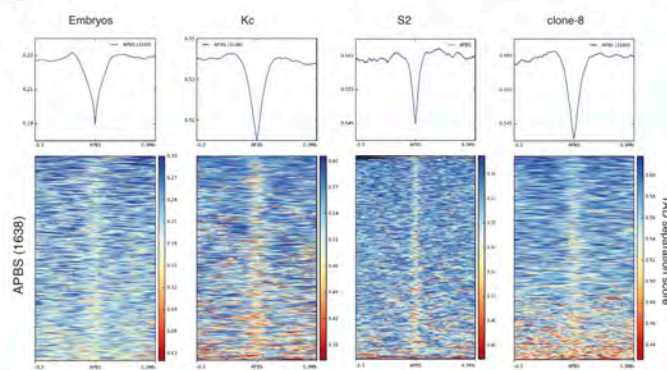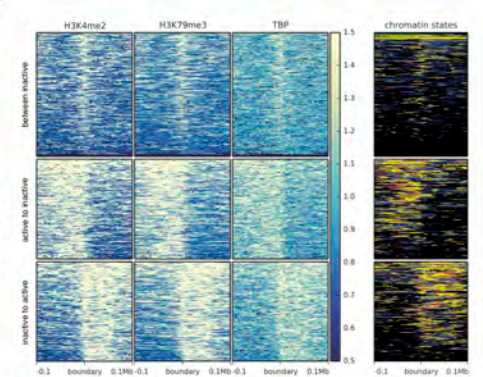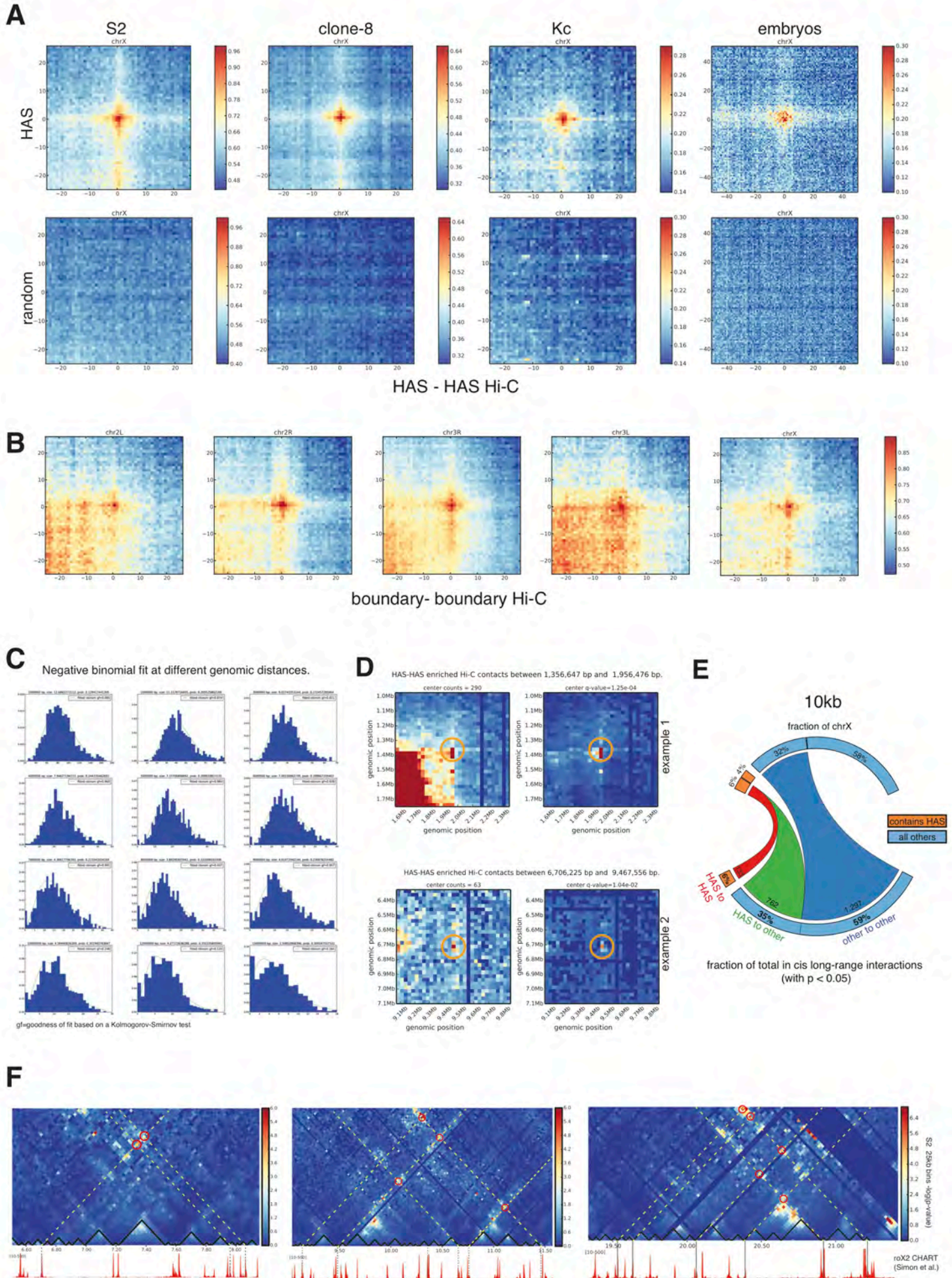vated) TAD boundary is conserved in all cases except in clone-8 cells. The TAD separation score from embryos shows more detail because of the higher resolution of this dataset compared to the other Hi-C. This is because the embryos' Hi-C was generated using a restriction enzyme that recognizes a motif of length four, found on average every 258 bp. In contrast, for the other Hi-C experiments the restriction enzyme used (HindIII) recognizes a motif of length 6 that is found on average every 1.720 bp. **E.** Fraction of common boundary positions for the different cell lines and for APBS. "embryo orig" and "kc orig" are the original

boundaries reported in (Sexton et al., 2012) and in (Hou et al., 2012) respectively. To compute the common boundaries we first increased the boundary to occupy a region of 20.000 bp, to account for the limited resolution of the Hi-C data, and then computed the number of overlaps between any two sets of boundaries over the size of the smallest set. **F.** Fraction of common boundary position for replicates. **G.** Comparison of distances from HAS to TAD boundaries using different methods showing that the definition of boundaries does not changes the overall result. The figures are like in Figure 1B. **H.** Heat maps showing the TAD separation score at architectural protein binding sites (ABPS) (Van Bortle et al., 2014). The center of the heat map is the APBS ±300 kb. **I.** Heat maps showing ChIP-on-chip profiles for H3K4me2, H3K79me3 and TBP DamID together with the chromatin states (on the right) (Filion et al., 2010). At the center of the heat maps is the boundary ±1 Mb. Using a k-means clustering based on the underlying heat map values, we identify three groups of boundaries: those that are between inactive regions, and those that mark the transition from active to inactive or vice versa. In the chromatin states color coding, yellow and red represent active chromatin and blue and black inactive chromatin.

Figure S4

A

S2   clone-8   Kc   embryos

HAS

random

HAS - HAS Hi-C

B

chr2L   chr2R   chr3R   chr3L   chrX

boundary- boundary Hi-C

C

Negative binomial fit at different genomic distances.

gf=goodness of fit based on a Kolmogorov-Smirnov test

D

HAS-HAS enriched Hi-C contacts between 1,356,647 bp and 1,956,476 bp.

center counts = 290   center q-value=1.25e-04

example 1

HAS-HAS enriched Hi-C contacts between 6,706,225 bp and 9,467,556 bp.

center counts = 63   center q-value=1.04e-02

example 2

E

10kb

fraction of chrX

contains HAS
all others

fraction of total in cis long-range interactions
(with p < 0.05)

F

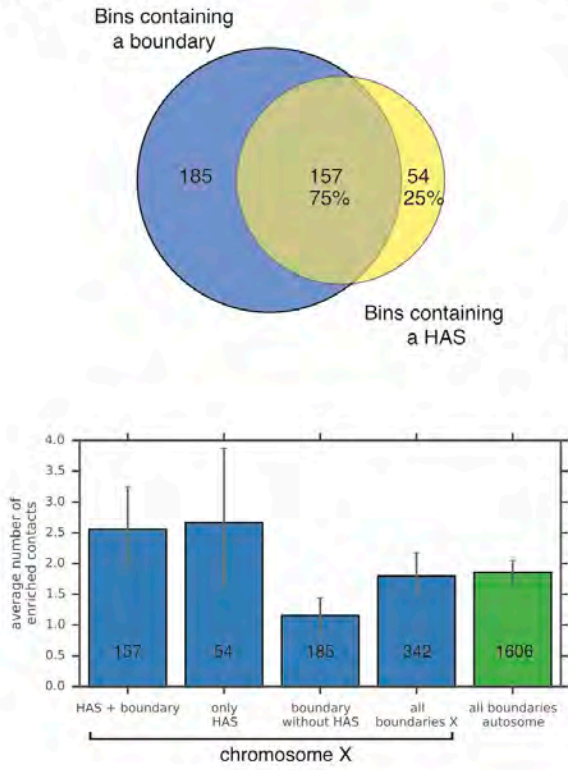
roX2 CHART
(Simon et al.)

S2 25kb bins -log(q-value)

Figure S4.

Enrichment of corrected Hi-C contacts between HAS loci and boundaries, related to Figure 2.
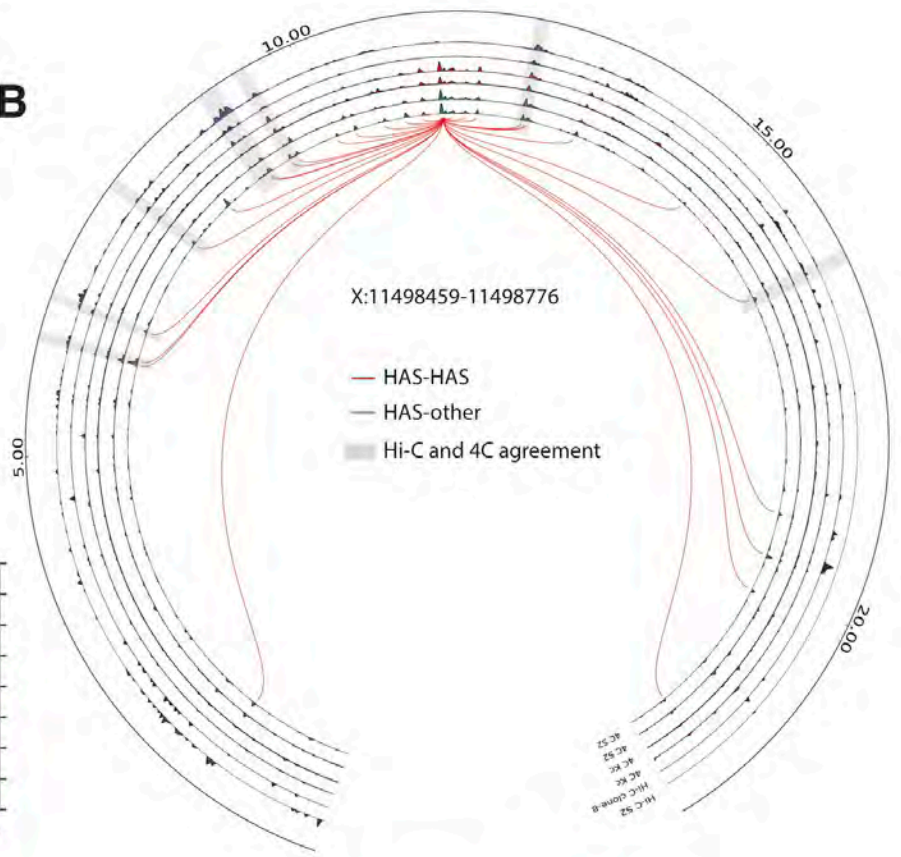
A. Similar to Figure 2A but in two dimensions. In the first row each heat map shows the HAS-HAS enrichment for the different cell lines and for embryos using restriction fragment resolution. The color scale represents the average number of contacts over many superimposing matrices (See Extended Experimental Procedures). In the axes, '0' indicates to the position of the HAS and the rest of the numbers represent the number of bins away from the HAS. The second row shows the shuffling of the HAS to a random position within the chromosome X. **B**. Enrichment of corrected Hi-C boundary-boundary contacts in the different chromosomes using S2 Hi-C data. **C.** Statistical estimation of enriched long-range contacts. Example of the negative binomial fit at different genomic distances, indicated on top of each panel. The fitting parameters size and probability are also written on top of each panel. The goodness of fit as measured using the Kolmogorov-Smirnov test is given on the legend. **D.** Two randomly selected examples of enriched contacts. The left panels contain the number of corrected Hi-C counts and the right panels contain –log(p-values) after multiple testing correction. **E.** Similar to Figure 2C but using 10 kb bins. To gain statistical power for 10 kb resolution we merged 8 Hi-C samples. As shown in the Figure 2C, HAS have a significant number of contacts between them and with other parts of the genome. **F.** –log(p-values), corrected for multiple testing, highlighting enriched long-range contacts between HAS for different genomic region on chromosome X. The lower track shows the roX2 CHART (Simon et al., 2011) in which peaks correspond to HAS. Red circles highlight HAS-HAS enriched contacts. Near TADs (represented by black triangles) our method is not able to detect enriched contacts, however, as seen for the 4C data, such contacts are present.

Figure S5

A

Bins containing a boundary

185   157 75%   54 25%

Bins containing a HAS

B

X:11498459-11498776

HAS-HAS
HAS-other
Hi-C and 4C agreement

C
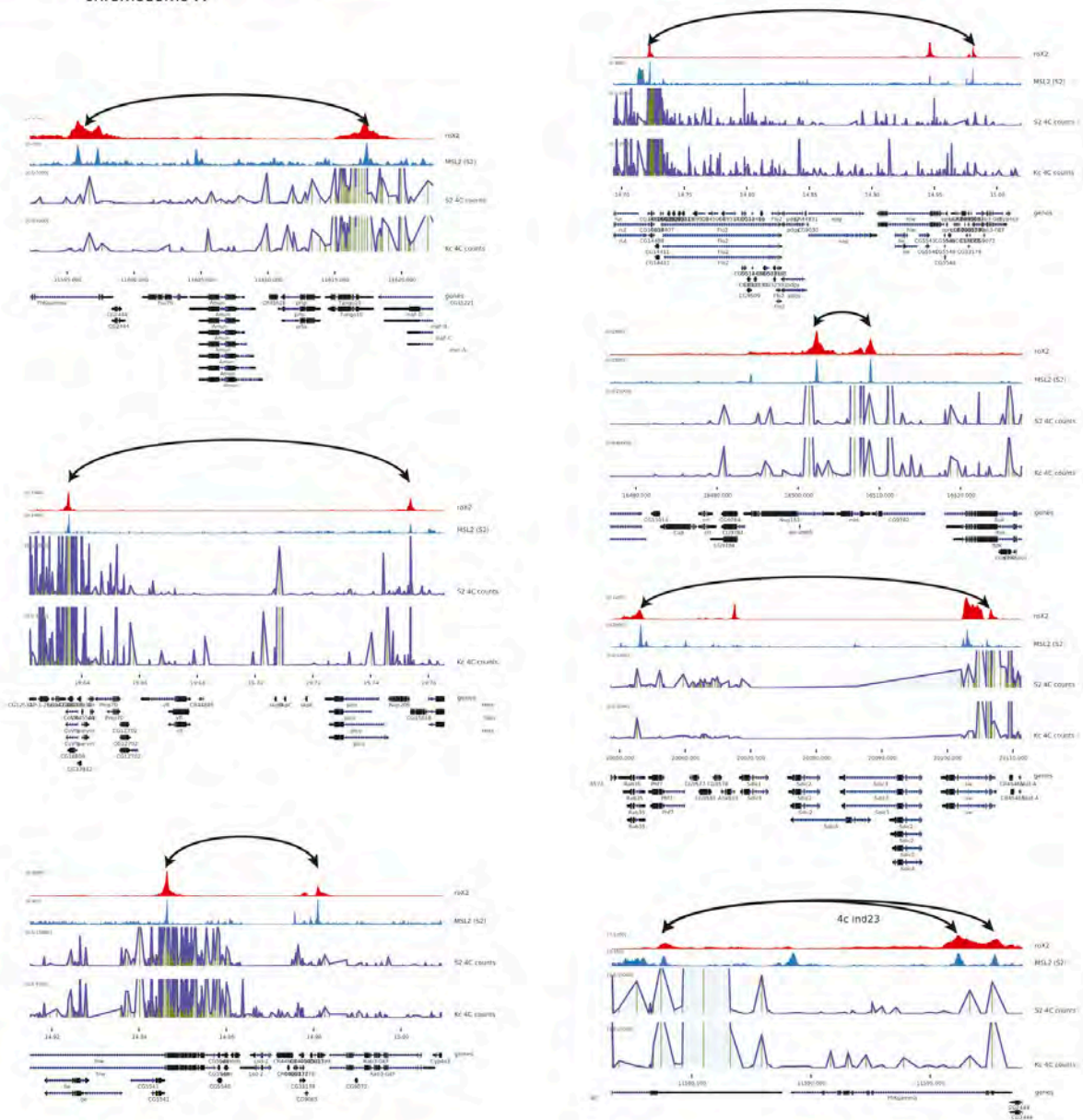
Figure S5.

HAS-HAS interactions, related to Figure 2.

A. (Top) Venn diagram showing the number of 25 kb bins that are boundaries or that contain a HAS. (Bottom) average number of contacts for different categories of bins. The 95% confidence intervals are estimated using bootstrapping. Bins that are not associated to boundaries tend to have fewer number of interactions, while bins that have a HAS have more long-range contacts. At 10 kb resolution the plot is very similar. **B.** Similar to Figure 2D showing a second example of the correlation of enriched contacts determined using either 4-C or Hi-C. **C.** HAS-HAS 4C contacts. Multiple examples of enriched 4C contacts between HAS (see also Figure 2E).

Figure S6

A

MSL1
MSL2
MOF
MSL3
tubulin

Wild type | MSL2 RNAi | EGFP RNAi | MSL3 RNAi

B

C

MSL2 RNAi r1 — MSL2 RNAi r2
MSL3 RNAi r1 — MSL3 RNAi r2
EGFP RNAi r1 — EGFP RNAi r2

D

Wild type
MSL2 RNAi
MSL3 RNAi
EGFP RNAi

chromosome X 14.000.000-15.500.000

E

4C contacts (S2)

replica 1

0 Mbp   5 Mbp   roX1   roX2   10 Mbp   2.6 Mbp insertion   15 Mbp   20 Mbp

replica 2

0 Mbp   5 Mbp   roX1   roX2   10 Mbp   2.6 Mbp insertion   15 Mbp   20 Mbp

F

Hi-C contacts (S2)

chr3L insertion

Hi-C corrected distances

0 Mbp   roX1   5 Mbp   10 Mbp   roX2   15 Mbp   20 Mbp   25 Mbp

Figure S6.

Hi-C for MSL2 and MSL3 RNAi, and distant 4C contacts from *roX2* locus, related to Figure 3.

A. MSL complex members were efficiently depleted in knockdown Hi-C samples compared to control. Wild type S2 cells and S2 cells treated with dsRNA targeting EGFP (control), MSL2 or MSL3 were lysed and subjected to western blot analysis with the indicated antibodies against the MSL complex members and Tubulin. Tubulin was used as a loading control. **B.** Similar to Figure S1E. The correlations were based on Hi-C counts using 10 kb bins. No significant differences are seen between wild type and knockdown Hi-C contact maps. **C.** Similar to Figure S4A. The two columns contain the replicate experiments while the rows show the MSL2 and MSL3 knock down and the EGFP RNAi control. No differences are observed in the knock downs with respect to the wild type or control samples. **D.** TAD structure and contact frequency for wild type, MSL knockdowns and EGFP RNAi control over an example region in chromosome X. **E.** The 4C data confirms the results obtained using Hi-C in Figure 3D showing that *roX2* contacts are biased towards the left side. Because the 4C computation of long-range distances does not take genomic distance into account, the 4C results for S2 show that this bias is not caused by an artifact on the computation of long-range contacts. **F**. Enriched Hi-C contacts after adding the 2.6 Mb chr3L insertion to the chromosome X showing the same bias in S2 cells.

Figure S7
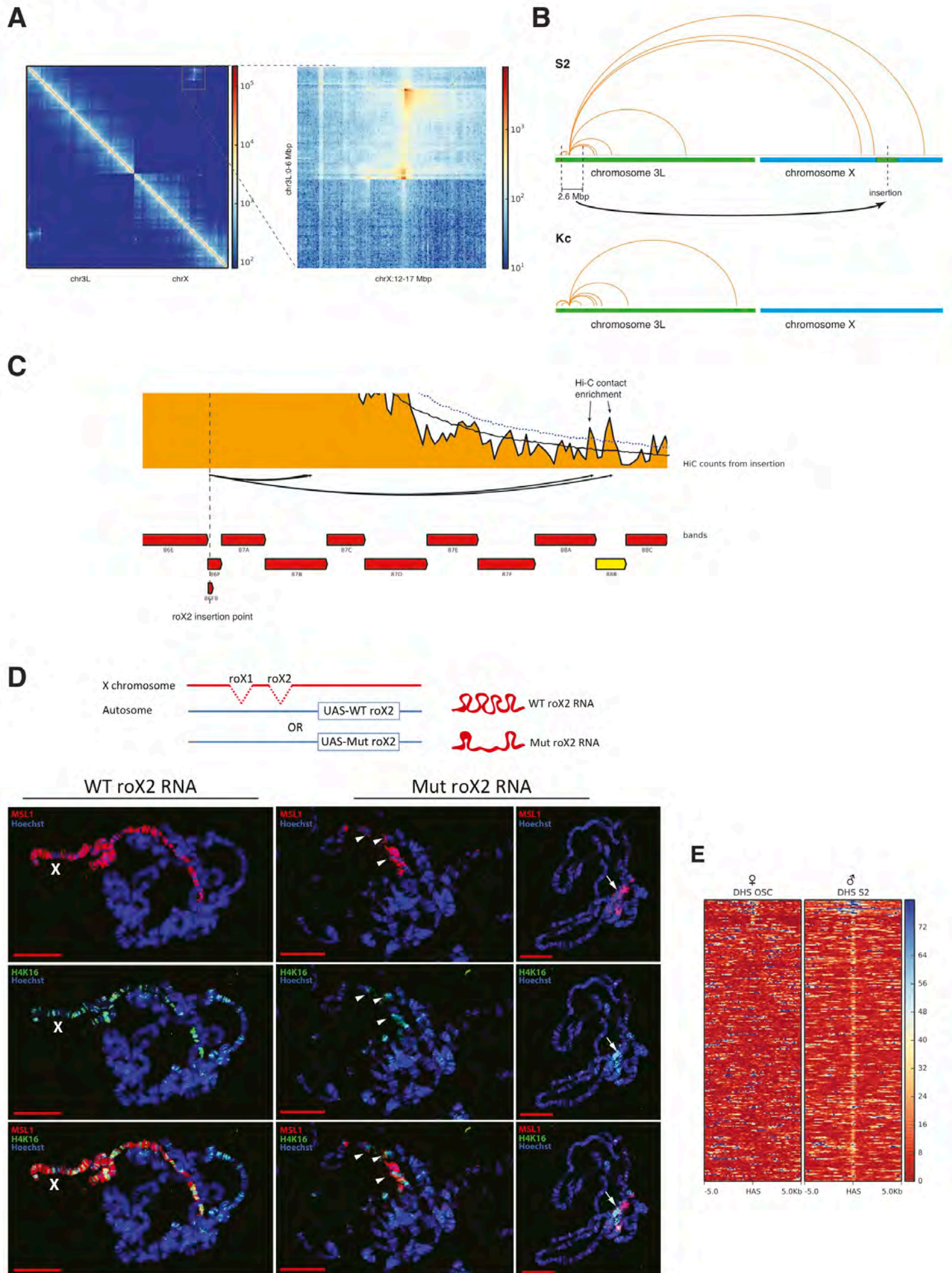
Figure S7.

Translocation from an autosome to X in S2 cells, expression of mutant roX from an autosome and DNAse hypersensitivity around HAS, related to Figures 5, 6 and 7.

A. Hi-C contact matrix reveals a translocation from chromosome 3L between positions 796,745 and 3,468,912 into chromosome X at around position 14,809,484. This region has 5 copies (Zhang et al., 2010), but only 1 or 2 are in chromosome X while the rest are not translocated. **B.** 4C viewpoint at position 1,795,000 bp inside the translocated region. For each plot the top 10 interactions are shown based on *p*-values. In S2 cells, the viewpoint has contacts with chromosome X whereas in Kc cells that do not have the translocation only the contacts within chromosome 3L are maintained. **C.** S2 Hi-C read counts (25 kb bins) from the site corresponding to the ectopic insertion of LacO-roX2 to band 88B. The black continuous line represents the chromosomal median counts for any two bins at that distance. The dotted line represents the 75 percentile**.** The image shows enriched contacts between the ectopic insertion of *roX2* at 86F8 and the MSL1 stained band 88B (yellow box) as seen in polytene chromosomal immunostainings (Figure 5B). **D.** Expression of a mutant form of roX RNA lacking important stem loop structures leads to mistargeting and male lethality. The autosomal insertion of a mutant form of roX2 RNA causes inefficient targeting of the MSL complex (marked by MSL1, red) to the X chromosome and mislocalization of H4K16 acetylation (in green) as visualized on polytene chromosomal immunostainings of salivary glands from male third instar larvae. DNA is stained with Hoechst 33342 (blue). All transgenes are inserted in the same autosomal location VK33 (See Extended Experimental Procedures for details). A schematic representation of the transgenes used is shown above. The label "X" indicates the X chromosome. Arrow heads point at impaired MSL targeting on X chromosome, while arrows point at mislocalization of MSL1 to the chromocenter. **E.** DNAse hypersensitivity (Arnold et al., 2013) for female OSC and male S2 *Drosophila* cell lines (GEO: GSE40739) around HAS.

Supplementary Table legends

Table S1. Processing of Hi-C sequencing reads, related to Figure 1.

This table contains mapping and filtering data for all Hi-C sequencing reads from S2, clone-8, Kc (Hou et al., 2012) and mixed embryos (Sexton et al., 2012).

Table S2. High-affinity sites, related to Figure 1.

This table contains the genomic coordinates of HAS determined using high-resolution roX CHART (Simon et al., 2011) and MSL2 ChIP-seq (Straub et al., 2013) in BED format.

Table S3. Interactions within HAS, from HAS to other regions and from other to other, related to Figure 2.

The table contains the underlying data for Figure 2C in the firsts two rows. These results are based on the computation of long-range contacts in S2 cells using a *p*-value cut-off of 0.05. The next two columns contain the results when the more stringent *p*-value cut-off of 0.01 is used. The following columns contain the results for clone-8, Kc and embryos and for the MSL knockdowns and the EGFP RNAi control in S2 cells. The next rows contain the results for 10 kb bins using a merge of all 8 Hi-C samples for S2 cells. For comparison, results obtained after shuffling the position of the HAS in the X chromosome and for boundaries in chromosomes 3L and 3R are also included in the remaining columns. In the table, the first column list the number of bins with enriched contacts, either HAS, random or boundaries. The fraction over the total number of bins appears below the counts. The second column contains the enriched contacts for all other bins. Columns three and four contain the number of bins without enriched contacts in the respective categories and the last three columns

contain the number of enriched contacts, and fraction, with respect to total number of contacts.

Table S4. 4C viewpoints, related to Figure 2 and 3.

This table contains, among other information, the location of the probe, the primers used and the number of sequencing reads.

Table S5. Differential analysis of DNA FISH data in clone-8 and S2R$^+$ cells, related to Figure 4.

Upper tables show number of possible permutations representing different scenarios when *roX1* and *roX2* contextually fall in different bins for their distance to the center of mass of the X chromosome territory (as defined in Figure 4B). Lower tables point at the extreme cases when only *roX1* and/or *roX2* are outside the territory (see Experimental Procedures for full description of the analysis).

Extended Experimental Procedures

Tissue culture and formaldehyde crosslinking

S2 cells were cultured in Schneider's medium (Gibco) supplemented with 10% FCS and 0.05% Pluronic F-68 (Sigma-Aldrich) on shaking incubators at 27°C at a density of 2-16 million/ml. Clone-8 cells were grown in Shields and Sang M3 Insect Medium (Sigma-Aldrich) supplemented with 2% FCS, 5μg/ml insulin and 2.5% fly extract on 15cm cell culture dishes. For crosslinking, 50 million S2 cells were collected by centrifugation and resuspended in 22.5ml of fresh medium in a 50ml falcon tube. Clone-8 cells were washed once with RT PBS, then 22.5ml of fresh medium were added. To obtain a final concentration of 1% FA, 0.625ml of ~37% stabilized FA solution was added and incubated at room temperature (RT) on a rotating wheel with 20rpm for S2 cells and a rocking platform in case of clone-8 cells. After 10 min 1.25ml of a 2.5M glycine solution was added to quench the FA and the cells incubated as before for additional 5 min. Subsequently, the S2 suspension or the clone-8 cell containing plates were placed on ice for 10 min. Clone-8 cells were then scraped off the plates into a 50ml falcon. In either case, the cell suspension was spun down in a 4°C cold centrifuge at 500g for 5 min, the supernatant removed, cell pellets frozen by incubation on dry ice for 20 min or snap-freezing in liquid nitrogen and then stored at -80°C until further processing.

RNAi in S2 cells and western blot analysis

dsRNA against EGFP, msl-2 and msl-3 was generated using Ribomax® Large Scale T7 in vitro transcription system (Promega) according to manufacturer's instructions using EGFP and Msl2 primers from (Larschan et al., 2011) and T7 Msl3 primers:

Msl3_fwd: taatacgactcactatagggGTCTGCATGCTGAAGGAGGT

Msl3_rev: taatacgactcactatagggTGTCACCGCTGGGATTGATA

designed with E-RNAi (Horn and Boutros, 2010) and purified with MEGAclear™(Ambion) and eluted in nuclease-free water to a concentration of 250ng/μl, 100mM NaCl and heated to 70°C for 10 min and left to cool to room

temperature. For knockdown experiments, 1 million cells were plated on 6-well dishes and after 12h washed with PBS and incubated with 1ml of serum-free Schneider's medium supplemented with 15μg of dsRNA. After one hour, 1ml of Schneider's medium +20%FCS was added and cells incubated at 25°C. After four days of knockdown, cells were collected by resuspension and split into aliquots for formaldehyde fixation or whole cell western blot analysis using previously described polyclonal antibodies against MSL1 and MSL3 raised in house in rats (Mendjan et al., 2006) and against MOF in rabbits (Conrad et al., 2012). MSL2 (d-300) antibody was purchased from Santa Cruz and α-tubulin antibody (EP1332Y) from Millipore.

Hi-C experimental procedure

Hi-C in S2 or clone-8 cells using HindIII as restriction enzyme was carried out as described by Belton et al. (Belton et al., 2012) with the following minor modifications: Starting material for all samples was 50 million insect cells per sample. After lysis, samples were taken up in 125μl and split into two aliquots of 50μl (Hi-C samples) and the remaining 25μl (3C control) were used to adjust for the smaller size of the *Drosophila* genome as compared to mammalian cells. Accordingly, for each 3C control only half the volumes per tube compared to the original protocol were used.

For the quality control of Hi-C libraries, the following primers were used:

fill-in_test_fwd CCTCCTAAAAGGGCATTCGAAATGAGCTGC

fill-in_test_rev: GATGCCCATTTGCCCGAAAAGTCCAAATTCC

After sonication, size selection was performed for ~230 bp DNA fragments and Illumina paired-end adapters (S2) or TruSeq adapters (clone-8) were attached. The final libraries were generated by 10 cycles of amplification with PfuUltra II Fusion DNA polymerase (Stratagene) for S2 samples or Q5® DNA polymerase (NEB) for clone-8 samples using Illumina PE Primer 1.0 and 2.0 (S2 cells) or barcoded TruSeq primers 1.0 and 2.0 (clone-8 cells) according to manufacturer's instructions. All libraries were sequenced on an Illumina HiSeq2000 DNA sequencer. Per sample, 150-300 million paired-end reads (2x50 bp) were generated. We produced two biological replicates per sample.

Hi-C processing

Each read mate was mapped individually with Bowtie2 (Langmead and Salzberg, 2012) using the --local option that trims the ends of a read if this increases the alignment score. This is useful to map Hi-C data because a significant portion of the reads contain a part of the ligated fragment (Imakaev et al., 2012) and can not align end-to-end to the reference genome. All reads that mapped multiple times to the genome were discarded as well as read pairs having a quality score of less than 20. Mapped read mates were then kept only if they fulfilled the following guidelines adapted from Jin et al. (Jin et al., 2013): *i.* each read must be within 800 bp of the nearest restriction site; this is to reflect the size selection prior to sequencing, *ii.* a mate pair of reads should not be duplicated (i.e. there is no other pair with the same mapping coordinates), *iii.* none of the mates starts with the restriction sequence as these are considered un-ligated reads, referred to as dangling ends, *iv.* 'inward' mates (reads facing each other) are only kept if they are over 1000 bp apart; all other mates are considered undigested fragments, *v.* 'outward' reads are kept only if they are 25 kb apart; these mates are considered to come from circularization of a fragment caused by self-ligation. For the Hi-C analysis in S2 cells, we obtained around 280 million reads per replicate. For the Hi-C analysis in clone-8 cells, we obtained 148 and 293 million reads for the two replicates, respectively (further details are found in Table S1).

We created two types of Hi-C matrices, one containing bins equal to the restriction fragment size and another of 5 kb bins. The resulting Hi-C matrices for 5 kb bins were merged to generate matrices of 10 kb, 25 kb, 50 kb and 100 kb. Replicates were merged into one dataset after evaluating their correlation. The data was normalized using the iterative correction method proposed by Imakaev et al. (Imakaev et al., 2012). Chromosomes were corrected individually, except when evaluating cell line translocations, to avoid biases stemming from different chromosome numbers (as in

the case of male cell lines). In this correction per chromosome we excluded inter-chromosomal interactions.

TAD separation score and boundary detection

Our TAD boundary detection is based on a simple graph clustering measure called the *conductance of a cut* (Kannan et al., 2004) that scores the interactions between two clusters with respect to the interactions within the clusters and is similar to other published measures (Sofueva et al., 2013; Van Bortle et al., 2014). The measure allows the identification of good clustering solutions characterized by few interactions between clusters. In our setting we consider Hi-C genomic bins as nodes in a graph, Hi-C counts as weighted edges, TADs as graph clusters of consecutive nodes (bins) and domain boundaries as genomic positions that best separate two clusters (TADs). To this end we quantify the strength of a TAD boundary by comparing the normalized Hi-C counts between the regions left and right of the boundary to the counts within such regions. For each bin we evaluate the *TAD separation score* between the left and right regions $(S, \overline{S})$ as follows:

$$\phi(S) = \frac{\sum_{i \in S; j \in \overline{S}} a_{i,j}}{\min\left(a(S), a(\overline{S})\right)}$$

The numerator is the sum of all contacts between the two clusters and $a(X)$ is the total sum of the contacts from X to any node on the graph V:

$$a(X) = \sum_{i \in X} \sum_{j \in V} a_{i,j}$$

To detect boundaries we first associate the corrected symmetric Hi-C matrix A to a graph with entries $a_{i,j}$ and then compute the TAD separation score, $\phi(i)$ for each cut separating bin i from bin $i+1$. Here, S contains the n consecutive bins to the left of the

cut and $\overline{S}$ the n consecutive bins to the right, where n is a parameter of the algorithm. S and $\overline{S}$ only contain bins in the same chromosome arm. To assure that S and $\overline{S}$ have always the same size, the measure is not computed for the n bins at the chromosome ends. A visual representation of a cut is given in Figure S3A. The (2) and (3) triangles contain all the Hi-C counts within the S and $\overline{S}$ clusters respectively. The box labelled (1) contains all edges between S and $\overline{S}$. All counts outside the larger triangle are not used. For cuts containing missing data the conductance $\phi(i)$ was defined as the average between $\phi(i+1)$ and $\phi(i-1)$. Boundaries are defined as positions where the conductance reaches local minima, $\phi(i) > \phi(i) < \phi(i+1)$. To distinguish pronounced local minima of conductance from noise, we also require that $\max[\phi(k)] - \phi(i) > d$, where the maximum is over the 3 nearest bins to the left and to the right of the cut. All local minima below a depth threshold d are discarded.

To determine the optimal parameters n and d of the algorithm, we used APBS binding sites (Van Bortle et al., 2014) as a positive set of true boundaries and inactive ('black') chromatin regions from Filion et al. (Filion et al., 2010) as a negative set. We found that the best parameters for the algorithm were for n=60 kb and d=0.05. The boundary detection resolution is related to the Hi-C resolution and the most precise boundary definitions are based on the embryos data from Sexton et al. (Sexton et al., 2012) in which a restriction enzyme (DpnII) that recognizes a four letter motif was used. The optimal values of n and d were similar for all cases evaluated.

In contrast to previously published methods to compute boundaries (Dixon et al., 2012; Hou et al., 2012; Sexton et al., 2012) our method is extremely simple and fast while producing comparable results (Figure S3C-G).

Enrichment of HAS at TAD boundaries
We used BEDtools (Quinlan, 2014) fisher method to determine the enrichment of HAS at TAD boundaries. Boundaries were extended to 5 kb left and right from the single base pair location given by TAD calling algorithms.

Enrichment of Hi-C contacts between HAS or boundaries

We used a method similar to the paired-end spatial chromatin analysis (PE-SCAn) (de Wit et al., 2013). For any pair of small genomic regions (e.g. HAS, boundaries, random positions) we selected a Hi-C submatrix centered at the intersection of the two regions (only upper half of matrix considered). We used matrices of corrected Hi-C counts at restriction fragment resolution, whose bins have a median length of 3.500 bp. All possible HAS intersections (5.979) within a genomic distance of 0.5 Mb to 15 Mb were considered. The submatrices selected contained 25 bins away from the center in both directions. Submatrices were then pooled together by computing their average at every position. To account for the fact that any submatrix that is closer to the main diagonal has higher counts compared to submatrices away from the main diagonal we normalized the matrix counts by dividing the values of the submatrix by the total number of Hi-C counts in the submatrix.

Estimation of significant long-range Hi-C contacts

To determine long-range Hi-C contacts we used a method similar to (Hou et al., 2012) and (Jin et al., 2013). This method allows the detection of significant enrichments of Hi-C counts based on the computation of background distributions stratified by genomic distance. The method is unrelated to the determination of TADs. We group all corrected Hi-C counts between bins that are at the same genomic distance and for each group we fit a negative-binomial distribution (Figure 6A) using the MASS package (Venables and Ripley, 2002) for R (only the upper-triangle of the symmetric matrix was used). Prior to fitting, outliers are removed using the median absolute deviation method (MAD) (Venables and Ripley, 2002). Subsequently, we test the goodness of fit between the data and the fitted distribution using the Kolmogorov-Smirnov and discard all cases with a poor fit (p-value < 0.01), which happens around 1-2% of the time. Next, each corrected Hi-C count in the matrix is evaluated using the respective distribution specific for the genomic distance between the row and column of the count. Finally, we estimate the false discovery rate (FDR) for all computed p-

values using the Benjamini & Hochberg method. To call long-range interactions we use a q-value < 0.05. To avoid potentially spurious long-range interactions from the occasional amplifications of the original Hi-C counts after iterative correction, we ignored the enrichments based on corrected Hi-C counts that were four or more times as large with respect the un-corrected counts. Furthermore, for our 25 kb bins analysis we required that in the original count matrix have at least 30 counts. We estimated long-range interactions per chromosome to avoid any spurious result caused by different ploidy of chromosomes, especially in male cells. To control for chromatin structure at the TAD range we ignore for the analysis any interaction < 500 kb. We center our analysis on 25 kb bins because they offer the more robust estimations. Smaller bins have fewer counts per entry and this reduces the statistical power for the detection of enrichments. Also, for matrices of 10 kb a significant number of bins have to be discarded either because no restriction enzyme site was found within the bin or because it contains few reads, usually because the restriction enzyme does not always cuts where a motif if found. For 10 kb bins around 15-30% of the total bins have to be discarded. However, to provide a robust analysis using 10 kb bins (Figure 6C) we merged together all Hi-C data from S2 cells: two wild type replicates, MSL2 and MSL3 knock downs and EGFP RNAi control that were largely identical. This allowed us to create a denser Hi-C contact matrix from about 540 million filtered reads.

4C-seq experimental procedure

4C-seq in S2 and Kc cells was carried out as described by Splinter et al. (Splinter et al., 2012) with minor modifications as follows: 50-100 million S2 cells or Kc cells, fixed as described above, were used for two biological replicates. DpnII (NEB) was used as the primary and Csp6I (ThermoScientific) as the secondary restriction enzyme. For each viewpoint, two 50µl PCR reactions (8 cycles with 55°C annealing temperature, followed by 18 cycles with 63°C) with 160ng of template each were prepared and the combined products cleaned up by a one-step purification using 1.6 volumes of AMPure® XP magnetic beads (Beckmann Coulter) according to

manufacturer's instructions. These viewpoint libraries were eluted in 20µl elution buffer (10mM Tris-HCl pH 8.5) and the molarities for each individual sample were estimated using the bioanalyzer DNA 7500 kit (Agilent Technologies). All different viewpoint libraries for one biological replicate were mixed equimolarly and sequenced on separate lanes on an Illumina HiSeq2500 DNA sequencer. Primer sequences and coordinates for the experiments are found in Table S4.

4C-seq bioinformatics analysis

The first mates from the paired-end sequencing were mapped using Bowtie2 (Langmead and Salzberg, 2012); bait sequences where trimmed. For the analysis of the vicinity interactions we used FourCSeq (Klein et al., 2014). For the analysis of global interactions we computed read coverage for bins of 100 bp using deepTools (Ramírez et al., 2014). Estimation of p-values was based on the method from Splinter et al. (Splinter et al., 2012) with the following modifications: The original method estimates the relative enrichment of the number of fragment ends covered by one or more reads (number of hits) in a small running window with respect to a background containing the number of hits over a larger window (both windows centered at the position being evaluated). The size of the small window is such that it contains 100 fragment ends and the large window size is chosen to contain 3,000 fragment ends. In our analysis, we instead use windows of fixed size, the small window having a size 20,000 bp and the large window with size 600,000 bp. These window sizes contain on average 100 and 3,000 fragment ends, respectively.

Comparison of 4C and Hi-C

To compare 4C and Hi-C, we retrieved the adjusted p-values from the row containing the respective 4C view from the Hi-C matrices. Such p-values represent the contact enrichment from the viewpoint to all other regions of the genome. These p-values were compared with the p-values obtained from the 4C-seq experiments.

Fly culture and genetics

Flies (*D. melanogaster*) were reared on a standard *Drosophila* medium at 25°C and 70% relative humidity on a 12h dark/12h light cycle. *w*;* The two transgenic lines carrying full-length *roX1* or *roX2* downstream of *UAS*, *w*;; P{UAS-roX1}VK00033 and w*;; P{UAS-roX2}VK00033*, respectively were generated through ΦC31 integrase-mediated germline transformation as previously described (Groth et al., 2004) to avoid the influence of position effects on gene expression and to facilitate direct comparison upon phenotypic analysis. Plasmid DNA was injected into *y1 M{vas−int.Dm}ZH-2A w*; PBac{y+−attP−3B}VK00033* embryos (Bloomington stock #24871), that carried an *attP* docking site at position 65B2 on chromosome arm 3L (Venken et al., 2006) and a *Drosophila* codon-optimized ΦC31 integrase driven in the germline by the *vasa* promoter (Bischof et al., 2007).

Transgenic flies carrying an autosomal *roX2* allele downstream of lacO (lacO-roX2+p) were generated by integration of a 256-mer lac operon array followed by the *roX2* coding sequence downstream of its putative endogenous promoter [the genomic region between *roX2* (*CR32665*) and the last exon of the upstream *CG11695* gene] into *y¹ M{vas-int.Dm}ZH-2A w*; M{3xP3-RFP.attP}ZH-86Fb* embryos (Bloomington stock #24749) carrying an attP docking site at position 86F8 on chromosome arm 3R and codon-optimized ΦC31 integrase driven in the germline by the *vasa* promoter (Bischof et al., 2007).

The following flies were obtained from the Bloomington stock center: *w*; P{Hsp83-GFP.lacI}2* (stock #25376), *w¹¹¹⁸; Dp(1;3)DC067, PBac{DC067}VK00033* carrying the duplicated segment 3F1;3F4 (X:3668064..3776288) that includes *roX1*, in position 65B2 (stock #30243) and *w¹¹¹⁸; Dp(1;3)DC511, PBac{DC511}VK00033/TM6C, Sb1* duplicating *roX2* as part of the segment 10C5;10D2 (X:11451601;11549239) inserted in 65B2 (stock #33494). The *roX1^SMC17A, roX2^Δ; CyO, hsp83-roX1* flies were kindly donated by V. Meller.

All lines used in this study were generated by standard genetic crosses from the above listed fly stocks.

Analysis of male-specific lethality rescue with *roX2* autosomal alleles

To analyze the extent of MSL complex spreading and positioning of the autosomal rescue allele relative to the X chromosome territory, $roX1^{SMC17A}$, $roX2^{\Delta}$ virgin females were crossed to either w^*; $P\{Hsp83\text{-}GFP.lacI\}2$; $lacO\text{-}roX2\text{+}p$, w^{1118}; $Dp(1;3)DC067$, $PBac\{DC067\}VK00033$ or w^{1118}; $Dp(1;3)DC511$, $PBac\{DC511\}VK00033$ male flies and kept at 25°C for mating and egg-laying for 3-5 days. The same approach was used to cross $roX1^{SMC17A}$, $roX2^{\Delta}$;; $daGal4$ virgin females with w^*;; $P\{UAS\text{-}roX1\}VK00033$ and w^*;; $P\{UAS\text{-}roX2\}VK00033$ *males*. To analyze the rescue efficiency of the autosome-derived roX2, eclosion rate and total number of flies was determined for both males and females from seven independent crosses by counting adult flies daily for a period of 10-12 days after the start of eclosion. Imaging analysis of male progeny larvae was carried out as described below.

For gene-expression analysis, wandering 3rd instar larvae of the appropriate genotype were homogenized in TRIzol (Qiagen). Total RNA was extracted from these lysates using the Direct-zol kit (Zymo) as per manufacturer's instructions and subsequently reverse transcribed using Superscript III (Life Tech.) and random hexamers. After real time qPCR with primers specific for 10 genes up stream and 10 genes downstream the 86Fb insertion site on chromosome 3R, transcript abundances were calculated using qPCR and the $2^{-\Delta\Delta Ct}$ method, normalizing against PFK (phosphofructokinase).

Analysis of male-specific lethality and effect on spreading in mutants with altered roX2 structure

To determine the effect on MSL complex spreading and male viability upon ectopic expression of $UAS\text{-}roX2^*$, $roX1^{SMC17A}$, $roX2^{\Delta}$; CyO, $hsp83\text{-}roX1$; tubGal4/TM6BTb virgin females were crossed to A1B4 (roX mutant) males (described in Ilik et al., 2013). Viability was assessed as described in the previous paragraph.

Immunostaining of polytene chromosomes and *Drosophila* tissues

Polytene chromosomes from 3rd instar male larvae were squashed and stained as described previously (Raja et al., 2010), with the following modifications: Methanol-free formaldehyde (FA) (16%, ThermoScientific) and a reduced amount of acetic acid (25% rather than 50%) were used in order to achieve a better GFP signal in the immunofluorescence (IF). Salivary glands and imaginal discs were dissected out from 3rd instar wandering larvae and fixed in 3.8% FA in PBS1X for 20 min at RT. After a few washes (4X15 min) in Washing buffer (1%Triton-X 100 in PBS1X, PBST), tissues were blocked in Blocking buffer (1% Triton-X 100, 2% BSA, in PBS1X) overnight (ON) at 4°C and then incubated with primary Ab for a few days (3-4) at 4°C. After washing in Blocking buffer (at least) 4X10 min, samples were incubated in secondary Ab and Hoechst (all 1:500) for 1 hour and 30 min in darkness, at RT. Samples were then washed 4X15 min with PBST and mounted in Fluoromount on a microscopy slide. Brains were dissected out from 3rd instar wandering larvae and subsequently placed in cold larva head buffer (protease inhibitors and 0.05% NP-40 in PBS1X). Samples were then spun down for ~1s in a table centrifuge and fixed in PBS containing 4% Methanol-free FA in PBS1X at RT for 7 min. Brains were washed twice with 1M glycine in PBS for 10s and three times with pure ice-cold larval head buffer, and then placed on a poly-L-lysine coated microscopy slide to be squashed after placing a coverslip on top. Slides were snap frozen in liquid nitrogen and for the subsequent steps the same protocol as for the polytene chromosome spreads was followed. A polyclonal antibody raised against MSL1 in rat prepared in house (Mendjan et al., 2006) and a GFP antibody purchased from Torrey Pines Biolabs Inc (TP401) were used for these experiments. H4K16Ac antibody was purchased from Santa Cruz. Incubation with the appropriate secondary antibodies was coupled with Hoechst 33342 (Life Technologies) diluted 1:500.

Confocal microscopy and Image analysis

Three-dimensional image stacks were taken with Zeiss LSM780 confocal microscope equipped with a 63X Plan-Apochromat oil-immersion objective with a numerical aperture of 1.4. 3D measurements between objects were performed using NEMO software (Iannuccelli et al., 2010). Parameters of the program were optimised in order to get the best accordance between the original images of nuclei and FISH probes and the result of the computer recognition.

Analysis of the relative position of *roX1* and *roX2* genes with respect to X chromosome territory

DNA FISH probes were labelled using the FISH Tag™ DNA Multicolor Kit (Life Technologies). The X chromosome territory was labelled with an antibody raised against MSL1 in rat generated in house. Images were acquired as z-stacks to allow a 3D view of the cells, converted into Multi-Tiff file format for each individual channel, and uploaded on the NEMO database. In the first step of the analysis the channel for Hoechst is used to detect cell nuclei and create a map of the field of view under observation. In the following steps NEMO detects object in all the other channels depending on size/intensity of the signal parameters set by the user and taking into consideration resolution (zoom and frame size) of each image. Finally, for each cell, an output table is generated with measurements within objects pairs, either between center of mass and/or border of the different objects, % co-localization, angle. If necessary, the parameters of the program were adjusted separately for each image to obtain the most adequate results. Data represented in Figure 4B and Table S5 has been generated by looking at the measurements calculated for the distances between the center of mass of *roX1* or *roX2* and the X chromosome territory. It was also taken into account the existence of permutations between all the possible combinations of *roX1* and *roX2* spatial positioning in the predefined bins for their distance from the center of mass of the X chromosome territory in clone-8 vs. S2 cells (Table S5).

Computation of a spot escaping from the X chromosome territory

Distances between the center of mass of a chosen probe and the X chromosome territory were then used to calculate the number of cells where one or both probes were found outside the X chromosome territory. Escaping spots were considered such when the % of co-localization probe-X chromosome territory was equal to 0 and, the distance of the probe from the border of the X chromosome territory was higher than 0. Furthermore, within these distinct cases, for *roX1* and *roX2*, we computed in a table the reciprocal position of each probe when the other one was detected outside the X chromosome territory. Images were arranged with Adobe Illustrator.

Determination of high affinity sites (HAS)

We took advantage of the recently published roX2 DNA binding results (Simon et al., 2011) to provide us almost base pair level resolution of HAS. For this, we used MACS2 (Feng et al., 2012) to compute enrichments (peaks) for roX CHART data and for MSL2 ChIP-Seq (Straub et al., 2013). HAS were defined as those regions having a roX2 and a MSL2 peak. This produced list of 257 positions. Visual inspection in the genome browser confirmed that our HAS were accurate and correlated with other previous HAS definitions (Alekseyenko et al., 2008; Straub et al., 2008), with MSL complex member, with roX2 ChIRP (Chu et al., 2011) and roX1 dChIRP (Quinn et al., 2014) peaks.

roX-DNA binding data

Unaligned reads for roX2 CHART (Simon et al., 2011) (GEO: GSE31332), and roX2 ChIRP (Chu et al., 2011) (GEO: GSM833475) were downloaded and mapped using the same methods as ChIP-seq data. The read coverage over bins of 50 bp was computed using deepTools (Ramírez et al., 2014). roX1 dChIRP tracks for clone-8 (Quinn et al., 2014) cells was downloaded from GEO (GEO: GSE53020).

ChIP-seq data processing

The following ChIP-seq data was downloaded from the Gene Expression Omnibus (GEO): MSL1, MSL2, MSL3, MOF, MLE and H4K16ac ChIP-Seq (Straub et al., 2013) for S2 cells (GEO: GSE37864); CTCF, Beaf-32 and CP190 (Wood et al., 2011) (GEO: GSE30740); condensin I, condensin II, cohesin (Van Bortle et al., 2014) (GEO: GSE54529). We downloaded the unmapped ChIP-seq and input reads, aligned them using Bowtie2 against the *D. melanogaster* BDGP Release 5 (dm3) genome assembly and produced coverage and log2 ratios using deepTools (Ramírez et al., 2014).

Determination of architectural protein binding sites (APBS)

We used the list of APBS provided by Van Bortle et al. (Van Bortle et al., 2014). Only positions bound by at least 7 architectural proteins (1638 in total, 295 on the X) were used. The proteins considered in their study where CTCF, Beaf32, Su(Hw), CP190, Mod(mdg4), DREF, Chromator, L(3)mbt, TF3C, Rad21, CAPH2.

Gene expression analysis

RNA-seq for S2 cells from Zhang et al. (Zhang et al., 2010) was obtained from (GEO: GSE16344). To identify dosage compensated genes, we downloaded the unmapped reads from two wild type replicates and two Msl2 RNAi samples. We mapped the reads using STAR (Dobin et al., 2013) to the *Drosophila* BDGP Release 5 (dm3) genome. Read counts per gene were obtained using feature Counts (Liao et al., 2014) based on the *Drosophila* RefSeq annotations. Genes having less than 10 read counts were removed from the analysis. Gene fold change (wild-type / Msl2 RNAi) was computed using DESeq (Anders and Huber, 2010). Active genes were determined using the method from Hart el al. ((Hart et al., 2013)) using RPKM values computed by Zhang et al.

MRE occurrences and annotation

Sequences of 200 bp from the center of the roX peak at HAS were used to compute the MRE using the MEME suite (Bailey et al., 2009). To detect genome-wide occurrences of the MRE, we used a balanced score-threshold as defined in Rahmann et al. (Rahmann et al., 2003) and we computed the TRAP score using the TRAP package (Thomas-Chollier et al., 2011). *Drosophila* gene annotations for the dm3 assembly were downloaded from UCSC Genome Browser Data (Rosenbloom et al., 2015).

ChIP-seq for MOF in transgenic flies

For each replicate, 170 male larvae with *roX2* inserted into the ectopic 86F8 site were grinded to powder in mortar with liquid nitrogen, and dounced with pestle B in NE buffer (15mM Hepes pH 7.6, 10mM KCl, 5mM MgCL2, 0.1mM EDTA, 0.5mM EGTA, 350mM sucrose, 0.1% Tween, 1mM DTT, protease inhibitors). The nuclei were fixed in 1.8% formaldehyde for 20 min at RT. After quenching for 5 min at RT with 125mM glycine, nuclei were further dounced for 30 times. The nuclei were then washed 3x 5min in RIPA (25mM HEPES pH 7.6, 150mM NaCl, 1mM EDTA, 1% Triton-X 100, 0.1% SDS, 0.1% DOC, protease inhibitors). The nuclei were broken with a Branson 250 sonicator at duty cycle 40, intensity 3, and chromatin was sheared by a Covaris S200 sonicator at duty cycle 10, intensity 200 for 30 min. For immunoprecipitation, MOF antibody was added and incubated overnight at 4°C. Immunocomplexes were pulled down by protein A-Sepharose. The beads were washed by 4x RIPA buffer, 1x LiCl buffer (10mM Tris at pH 8, 0.25M LiCl, 0.5% NP-40, 0.5% DOC, 1mM EDTA), and 3x TE buffer. The beads were resuspended in TE buffer. The input chromatin and the beads were shaken at 65°C O/N. After 30 min incubation at 37°C with RNaseA (0.2 mg/ml), followed by 2 hr Proteinase K digestion (0.05 mg/ml) at 50°C, DNA was purified using MinElute columns (Qiagen). Pair-end sequencing libraries were prepared using NEBNext® ChIP-Seq library prep reagent kit, according to manufacturer's instruction, and sequenced using Illumina HiSeq 2500 machines. For each ChIP replica we obtained 36 and 40 million reads

respectively and for the matching inputs we obtained 35 and 41 million reads. Reads were aligned to the dm3 *Drosophila* assembly using Bowtie2 (Langmead and Salzberg, 2012). Aligned reads were then processed using deepTools (Ramírez et al., 2014) to compute log2 ratios for MOF ChIP/input at 10 bp bins.

Nucleosome positioning analysis in *Drosophila*

50 million S2 cells were harvested and washed with PBS. Same number of cells was collected for each sample. The cells were resuspended in 1ml of crosslinking solution (50mM HEPES, 100mM NaCl, 1mM EDTA, 0.5mM EGTA) and 37% formaldehyde was added to a final concentration of 1%. The tube was rotated at room temperature for 10 min and the reaction was quenched with 2.5M glycine at a final concentration of 125mM. The cells were spun down immediately after addition of glycine and the solution was replaced with fresh PBS that contained 125mM glycine. The cells were resuspended and rotated at room temperature for 5 min. The cells were pelleted into equal aliquots by centrifugation with 2000 rpm for 5 min. Aliquots were then snap frozen in liquid nitrogen.

Cells were permeabilized with NP-40. The chromatin was digested with microcccal nuclease (MNase) at 25ºC for 10 min. Digested chromatin was then analyzed on an agarose gel to reveal the digestion pattern. We aimed to achieve a slightly under-digested condition with di- and tri-nucleosomes still visible. The digestion pattern of each sample was exactly the same. The MNase digestion was stopped by addition of EDTA, NaCl and SDS to a final concentration of 20mM, 150mM and 1% respectively. The samples were placed on ice during handling to ensure that the MNase digestion was halted completely. The chromatin was then de-crosslinked overnight at 65ºC overnight in TE buffer. The sample was treated with RNaseA (0.2 mg/ml) for 30 min at 37ºC, and with Proteinase K (0.05 mg/ml) for 2 hours at 50ºC. Finally, DNA was purified using MinElute columns (Qiagen) and eluted in nuclease-free water. The eluted DNA was then analyzed with the E-Gel Pre-cast Agarose Gels system (Invitrogen). Bands at positions corresponding to 150 bp, 300 bp, 450 bp,

represented mono- di and tri-nucleosomes, respectively. The 150 bp fragment was purified from the E-gel according to manufacturers specifications and used to generate the library for the Illumina paired-end sequencing. To minimize the effect of PCR artifacts, only 3 cycles of PCR were performed in the library generation procedure.

Separation of nucleosomal and sub-nucleosomal fragments by non-negative matrix factorization

MNase-seq yields fragments that correspond to nucleosomal and sub-nuclesomal fragments (Henikoff et al., 2011). In order to separate these two types of fragments fro the MNase-seq data we used a method referred to as non-negative matrix factorization (NMF). NMF aims at decomposing observed signals into a specified number of additive source signals. The result of this operation is (i) the source signal and (ii) the mixing proportions. In our case we determined for each fragment length between 50 and 300 base pairs the total number of fragment centers in a window spanning +/- 2000 base pair around transcription start sites (TSS) of active genes. This yields a matrix, where the rows correspond to the base pair relative to the TSS and columns denoting the fragment length. This matrix served as input to NMF using the R-package NMF (Gaujoux and Seoighe, 2010), where we used the Brunet method (Brunet et al., 2004) and two source components. The NMF yields two distinct signals, one we attributed to sub-nucleosomal fragments and the other to nucleosomal fragments. We took the learned mixing coefficients, i.e. the contribution of each fragment length to the nucleosomal source signals, to determine the signal strength corresponding to the nucleosomal fragments, i.e. each fragment contributed to the nucleosomal-sized source signal corresponding to its learned mixing coefficient. We repeated this approach for each experiment.

Estimating nucleosome positioning strength

Each nucleosomal fragment has signal strength according to the NMF. We accumulated the signal strength of the 5' and 3' end of each fragment and determined

the normalized difference per position in the genome between the ends mapping on the plus and minus strand divided by the total number of ends. We expect that a well positioned nucleosome should have many more plus strand ends than minus strand ends on the 5' end of the nucleosomal DNA, while on the 3' end of the nucleosomal DNA, we expect many more minus strand ends than plus strand ends. Thus, we took the normalized difference between plus and minus strand ends and convolved it with a linear filter of length 147, which starts at +1 at position 1 and ends at -1 at position 147. The resulting values correspond to our measure of nucleosome positioning.

Statistical significance of nucleosome positioning changes

The statistical significance of the change in the nucleosome pattern around HAS upon MSL2 depletion was obtained by computing individual p-values for change of the nucleosome signal (wild type vs. knockdown) for each base pair using ANOVA. Using Fisher's method the multiple p-values were combined. Number of HAS=257.

Supplemental References

Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.-K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J., et al. (2008). A Sequence Motif within Chromatin Entry Sites Directs MSL Establishment on the Drosophila X Chromosome. *134*, 599–609.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol.

Arnold, C.D., Gerlach, D., Stelzer, C., Boryn, L.M., Rath, M., and Stark, A. (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science *339*, 1074–1077.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Belton, J.-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi–C: A comprehensive technique to capture the conformation of genomes. Methods *58*, 268–276.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized

transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Pnas *104*, 3312–3317.

Brunet, J.-P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Pnas *101*, 4164–4169.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Mol Cell *44*, 667–678.

Conrad, T., Cavalli, F.M.G., Holz, H., Hallacli, E., Kind, J., Ilik, I., Vaquerizas, J.M., Luscombe, N.M., and Akhtar, A. (2012). The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. Dev. Cell *22*, 610–624.

de Wit, E., Bouwman, B.A.M., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J.A.M., Krijger, P.H.L., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227–231.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc *7*, 1728–1740.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *143*, 212–224.

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

Hart, T., Komori, H.K., LaMere, S., Podshivalova, K., and Salomon, D.R. (2013). Finding the active genes in deep RNA-seq gene expression studies. BMC Genomics *14*, 778.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., Macalpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proc. Natl. Acad. Sci. U.S.a. *108*, 18318–18323.

Horn, T., and Boutros, M. (2010). E-RNAi: a web application for the multi-species design of RNAi reagents--2010 update. *38*, W332–W339.

Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. Mol Cell *48*, 471–484.

Iannuccelli, E., Mompart, F., Gellin, J., Lahbib-Mansais, Y., Yerle, M., and Boudier, T. (2010). NEMO: a tool for analyzing gene and chromosome territory distributions from 3D-FISH experiments. Bioinformatics *26*, 696–697.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Meth *9*, 999–1003.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature.

Kannan, R., Vempala, S., and Vetta, A. (2004). Journal of the ACM (JACM).

Klein, F.A., Anders, S., Pakozdi, T., Ghavi-Helm, Y., Furlong, E.E.M., and Huber, W. (2014). FourCSeq: Analysis of 4C sequencing data.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth *9*, 357–359.

Larschan, E., Bishop, E.P., Kharchenko, P.V., Core, L.J., Lis, J.T., Park, P.J., and Kuroda, M.I. (2011). X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature *471*, 115–118.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., et al. (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. Mol Cell *21*, 811–823.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics *47*, 11.12.1–11.12.34.

Quinn, J.J., Ilik, I.A., Qu, K., Georgiev, P., Chu, C., Akhtar, A., and Chang, H.Y. (2014). Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat Biotechnol *32*, 933–940.

Rahmann, S., Müller, T., and Vingron, M. (2003). On the power of profiles for transcription factor binding site detection. Stat Appl Genet Mol Biol *2*, Article7.

Raja, S.J., Charapitsa, I., Conrad, T., Vaquerizas, J.M., Gebhardt, P., Holz, H., Kadlec, J., Fraterman, S., Luscombe, N.M., and Akhtar, A. (2010). The nonspecific lethal complex is a transcriptional regulator in Drosophila. Mol Cell *38*, 827–841.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *42*, W187–W191.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. *43*, D670–D681.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *148*, 458–472.

Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. U.S.a. *108*, 20497–20502.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. Embo J *32*, 3119–3129.

Splinter, E., de Wit, E., van de Werken, H.J.G., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods *58*, 221–230.

Straub, T., Grimaud, C., Gilfillan, G.D., Mitterweger, A., and Becker, P.B. (2008). The Chromosomal High-Affinity Binding Sites for the Drosophila Dosage Compensation Complex. PLoS Genet *4*, e1000302.

Straub, T., Zabel, A., Gilfillan, G.D., Feller, C., and Becker, P.B. (2013). Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq. Genome Research *23*, 473–485.

Thomas-Chollier, M., Hufton, A., Heinig, M., O'Keeffe, S., Masri, N.E., Roider, H.G., Manke, T., and Vingron, M. (2011). Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nat Protoc *6*, 1860–1869.

Van Bortle, K., Nichols, M.H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z.S., and

Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol *15*, R82.

Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S (New York: Springer).

Wood, A.M., Van Bortle, K., Ramos, E., Takenaka, N., Rohrbaugh, M., Jones, B.C., Jones, K.C., and Corces, V.G. (2011). Mol Cell *44*, 29-38.

Zhang, Y., Malone, J.H., Powell, S.K., Periwal, V., Spana, E., Macalpine, D.M., and Oliver, B. (2010). Expression in Aneuploid Drosophila S2 Cells. PLoS Biol *8*, e1000320.